

Zur Berechnung von Informationsflüssen in nichtlinearen dynamischen Systemen

Vom Fachbereich Physik
der Technischen Universität Darmstadt

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigte Dissertation

von
Dipl.-Phys. Olaf Stefan Göhrs
aus
Göttingen



Darmstadt 2000

D 17

Referent:	Prof. Dr. Theo Tschudi
Korreferent:	Privatdozent Dr. Wolfram Just
Tag der Einreichung:	07.12.1999
Tag der Prüfung:	09.02.2000

Inhaltsverzeichnis

1	Einleitung	1
2	Zeitreihen, dynamische Systeme und Information	4
2.1	Zeitreihen	4
2.2	Dynamische Systeme	4
2.3	Chaotische Dynamik	5
2.4	Entropie und Information	6
2.5	Kolmogorov-Sinai-Entropie	8
3	Informationsverluste über einzelne Meßgrößen	10
3.1	Der eindimensionale Fall	10
3.2	Der mehrdimensionale Fall	12
3.3	Wahl der ϵ -Umgebung	14
3.4	Zusammenhang mit der Kolmogorov-Sinai-Entropie	17
4	Bestimmung der bedingten Entropie aus Zeitreihen	19
4.1	Abschätzung bedingter Entropien	19
4.2	Korrektur der nichtlinearen Einflüsse	21
4.3	Iteration des Verfahrens	24
4.4	Abhängigkeit der bedingten Entropie $\tilde{H}_\epsilon(Y X)$ von der Umgebungsgröße und der Länge der Zeitreihe	28
5	Abhängigkeiten innerhalb einer skalarwertigen Zeitreihe	34
5.1	Dichten entlang einer Trajektorie	34
5.2	Zeitentwicklung der Dichten bei iterierten Abbildungen auf einem Intervall	36
5.3	Bedingte Entropie und Informationsflüsse in skalaren Zeitreihen .	39
6	Rekonstruktion nichtlinearer Dynamik	42
6.1	Bedingte Entropie und die Wahl der Verzögerungszeiten	42
6.2	Variable Verzögerungszeiten und die Bestimmung der Einbettungsdimension	46

7	Raum-zeitliche Zusammenhänge	51
7.1	Gekoppelte Abbildungen	51
7.1.1	Räumliche Zusammenhänge	52
7.1.2	Raum-zeitliche Zusammenhänge	55
7.2	Getriebene Systeme	59
7.2.1	Der Einfluß der Phasenraumgeschwindigkeit k	62
7.2.2	Einfluß der Kopplungsamplitude c'	64
7.2.3	Ergebnisse	65
8	Bedingte Entropien in experimentellen Zeitreihen	66
8.1	Rückkopplungserscheinungen in Halbleiterlasern	66
8.1.1	Halbleiterlaser mit zwei externen Resonatoren	66
8.1.2	Zeitliche Abhängigkeiten	67
8.2	Quasiperiodisch getriebener, nichtlinearer Schwingkreis	70
9	Datengestützte Modellbildung	74
9.1	Parameterermittlung	74
9.2	Termauswahlverfahren	76
9.2.1	Analyse von Turbogeneratoren	77
9.3	Bedingte Entropie und Modellbildung	78
9.3.1	Zusammenhänge zwischen den Meßgrößen	78
9.3.2	Vorgehensweise bei der Modellbildung	80
9.4	Echtzeitanalyse	81
10	Andere Methoden zur Bestimmung von Abhängigkeiten	85
10.1	Der Einfluß von Meßrauschen	85
10.2	Methoden zur Bestimmung von Abhängigkeiten innerhalb von Zeitreihen	89
10.2.1	Bedingte und relative Entropie	89
10.2.2	Transinformation	90
10.2.3	Kovarianz	92
10.2.4	Zusammenfassung	93
11	Zusammenfassung	95

A	Wahrscheinlichkeitstheoretische Grundlagen	98
B	Wichtige Begriffe aus Informations- und Ergodentheorie	100
B.1	Ergodentheoretische Begriffe	100
B.2	Informationstheoretische Begriffe	100
C	Echtzeitanalyse – der NLyzer	102
C.1	Aufbau des NLyzers	102
C.2	Beispiel: Modellbildung bei Motorlüftern	103
D	Symbolverzeichnis	105
	Literaturverzeichnis	106

1 Einleitung

Bei der experimentellen Untersuchung physikalischer Systeme liegen erste Ergebnisse häufig in der Form von Zeitreihen vor. Es ist die Aufgabe der Zeitreihenanalyse, aus diesen Messungen relevante Informationen über das jeweilige System zu extrahieren. Eine wichtige Teilaufgabe ist dabei, Zusammenhänge zwischen verschiedenen Meßgrößen oder auch Informationsflüsse innerhalb einer Meßreihe zu identifizieren. Mit dieser Information ist dann häufig möglich, physikalische Effekte besser zu erklären oder auch ein dynamisches System zu modellieren.

Informationstheoretische Ansätze ermöglichen dabei, derartige Zusammenhänge zwischen verschiedenen Größen zu identifizieren und zu quantifizieren. Insbesondere die bedingte Entropie ist hervorragend geeignet, funktionale Abhängigkeiten zwischen Meßgrößen zu charakterisieren. Sie basiert auf der Shannon-Entropie $H(Y)$, welche angibt, wieviel zusätzliche Information nötig ist, um jeden möglichen Zustand einer Meßgröße Y eindeutig vorherzusagen. Zusätzliche Information ist durch die Kenntnis einer weiteren Meßgröße X gegeben. Die bedingte Entropie $H(Y|X)$ gibt darüber Auskunft, wieviel Unkenntnis über Y verbleibt, wenn X bereits bekannt ist. Sie hat folgende wichtige Eigenschaften:

- Hat X keinen Einfluß auf Y , so ist $H(Y|X) = H(Y)$.
- Je mehr Kenntnis über Y vorliegt, desto kleiner ist die bedingte Entropie.
- Ist Y vollständig durch X bestimmt, verschwindet die bedingte Entropie.

Die bedingte Entropie gibt folglich an, wieviel Information über eine physikalische Meßgröße in anderen Größen enthalten ist. In dieser Arbeit wird eine Methode vorgestellt, die ermöglicht, eine modifizierte bedingte Entropie aus Zeitreihen zu berechnen.

Bei der Anwendung dieses Verfahrens können für einen positiven Wert der bedingten Entropie verschiedene Ursachen vorliegen. Einerseits ist es möglich, daß nicht genügend Messungen vorliegen, um eine eindeutige Vorhersage zu treffen. Andererseits können zufällige Störungen des Systems zu Informationsverlusten führen. In kontinuierlichen Systemen müssen außerdem nichtlineare Einflüsse berücksichtigt werden.

Der Informationsverlust in nichtlinearen dynamischen Systemen wird durch die Kolmogorov-Sinai-Entropie angegeben (Kapitel 2). Häufig kann es aber auch von Interesse sein, wieviel Information über eine einzelne skalare Meßgröße eines mehrdimensionalen Systems gewonnen werden kann. In Kapitel 3 werden daher die der Kolmogorov-Sinai-Entropie zugrundeliegenden Ideen auf diesen Fall übertragen.

Unter Berücksichtigung dieser Überlegungen wird in Kapitel 4 eine Methode entwickelt, durch welche eine modifizierte bedingte Entropie berechnet werden kann, die auch im Fall nichtlinearer Abhängigkeiten verschwindet. Mit dieser Methode können einerseits Abhängigkeiten quantifiziert werden, andererseits bietet die Iteration des Verfahrens die Möglichkeit, die für eine Modellierung optimal geeigneten Meßgrößen zu bestimmen.

Neben Abhängigkeiten zwischen verschiedenen Meßgrößen können auch zeitliche Abhängigkeiten innerhalb einer skalaren Meßgröße berechnet werden. Dazu werden zeitversetzte Sequenzen einer einzelnen Zeitreihe untersucht (Kapitel 5). Die Eigenschaft der modifizierten bedingten Entropie, bei gleicher zusätzlicher Information nicht geringer zu werden, wird dazu verwendet, optimale Parameter für eine Zeitverzögerungseinbettung zu bestimmen (Kapitel 6).

Werden zeitversetzte Sequenzen verschiedener Meßgrößen untersucht, können raumzeitliche Abhängigkeiten identifiziert werden. Dies wird im zeitdiskreten Fall anhand von gekoppelten Abbildungen, im kontinuierlichen Fall anhand getriebener Systeme demonstriert (Kapitel 7).

Die Einsetzbarkeit der vorgestellten Methode bei experimentellen Daten wird in Kapitel 8 gezeigt. Dazu soll der Einfluß von Rückkopplungen in einem Halbleiterlaser mit zwei externen Resonatoren bestimmt werden. Weiterhin wird ein quasiperiodisch getriebener, elektrischer Schwingkreis untersucht.

Durch die Bestimmung der modifizierten bedingten Entropie aus multivariaten Zeitreihen können die für eine Modellierung optimal geeigneten Meßgrößen bestimmt werden. In Kapitel 9 wird eine erfolgreiche Vorgehensweise bei einer derartigen Modellbildung beschrieben. Dabei wird am Beispiel der Analyse von Turbogeneratoren der Vorteil einer vorhergehenden Entropieanalyse demonstriert. Weiterhin wird die Einsetzbarkeit der Modellbildung in ein Programmpaket zur Echtzeitanalyse diskutiert. Dieses Programmpaket ist im Rahmen dieser Arbeit

entwickelt worden (Anhang C).

Abschließend wird das Verhalten des vorgestellten Verfahrens bei Meßrauschen untersucht. In diesem Zusammenhang werden andere Verfahren zur Bestimmung von Zusammenhängen in Meßreihen beschrieben, sowie deren Vor- und Nachteile diskutiert (Kapitel 10).

2 Zeitreihen nichtlinearer dynamischer Systeme und ihr Informationsgehalt

Die zentrale Frage dieser Arbeit ist, wieviel Information über die Dynamik eines nichtlinearen Systems aus Zeitreihen von Messungen des Systems erhalten werden kann. In diesem einleitenden Kapitel werden daher die für die weitere Arbeit wesentlichen Begriffe aus den Bereichen Zeitreihenanalyse, dynamische Systeme und Informationstheorie vorgestellt.

2.1 Zeitreihen

Eine *Zeitreihe* beschreibt die zeitliche Abfolge von verschiedenen Zuständen. Damit ist sie mathematisch gesehen die Realisierung eines stochastischen Prozesses $(\Omega, \mathcal{A}, P, X_{t \in T})$ (Def. A.8), wobei Ω eine beliebige Menge, \mathcal{A} eine zugehörige σ -Algebra, P ein Wahrscheinlichkeitsmaß und T eine Indexmenge ist, welche die Zeit repräsentiert; in der Regel ist $T \subset \mathbb{R}_0^+$ oder $T = \mathbb{N}_0$. Zu jeder beliebigen Zeit nimmt die Zufallsvariable $X(\omega)$, $\omega \in \Omega$, einen bestimmten Wert an: $X(\omega) = x_\alpha$, bzw. $X(\omega) = x$. Durch den griechischen Index werden dabei diskrete, also abzählbare Zustände gekennzeichnet. Entfällt dieser, so sind die möglichen Realisierungen stetig. Mit streng monoton wachsender Zeit t kann somit eine Zeitreihe als

$$\{x_{(\alpha),t}\}_{t \leq T} = \{X_t(\omega) : \omega \in \Omega, t \leq T\} \quad (2.1)$$

definiert werden. Eine *Meßgröße* wird also durch eine Zufallsvariable X beschrieben, die Folge ihrer Meßwerte zu den Zeiten t bildet eine Zeitreihe.

Die Realisierungen des stochastischen Prozesses als Funktion der Zeit bezeichnen auch einen Pfad, bzw. eine Trajektorie im Zustandsraum [1].

Diese sehr allgemeine Definition enthält noch keinerlei Aussage über die Beschaffenheit der Zeitreihe. Die Frage, ob die Abfolge der verschiedenen Zustände einen kausalen Zusammenhang aufweist, bleibt unberücksichtigt.

2.2 Dynamische Systeme

Die zeitliche Entwicklung im Zustandsraum wird durch ein *dynamisches System* beschrieben. Dabei handelt es sich um eine Familie von Abbildungen $f^t : M \rightarrow M$

auf einem Teilraum M des Zustandsraumes, wobei der Parameter $t \geq 0$ wiederum die Zeit beschreibt.

Im Falle diskreter Zeit kann die Dynamik durch eine iterierte Abbildung beschrieben werden

$$\mathbf{x}_{t+1} = f^1(\mathbf{x}_t). \quad (2.2)$$

Für zeitkontinuierliche Systeme ist eine Darstellung mittels Differentialgleichungssystemen üblich

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}). \quad (2.3)$$

Der *Fluß* f^t ergibt sich dann als Lösung des Differentialgleichungssystems. Ist f^t stetig differenzierbar, so spricht man von einem differenzierbaren dynamischen System.

Für einen beliebigen Anfangspunkt $\mathbf{x}_0 \in M$ ergibt sich durch Anwendung der Entwicklungsgleichungen (2.2), bzw. (2.3) eine Trajektorie im Zustandsraum. Durch Abtastung in diskreten Zeitschritten hat diese auch im zweiten Fall die Form der oben definierten Zeitreihe (2.1).

Meßreihen und mittels Computer generierte Zeitreihen besitzen in der Regel nicht den durch ein dynamisches System beschriebenen rein deterministischen Charakter. Einerseits werden diese Zeitreihen von Meßrauschen überlagert; dieses ist ein additives Rauschen, das die eigentliche Dynamik nicht beeinflußt. Andererseits kann auch ein dynamisches Rauschen vorliegen. Dabei handelt es sich um ein multiplikatives Rauschen; jede Störung des Systems verändert dessen weitere zeitliche Entwicklung. Bei numerisch erzeugten Daten liegen diese Störungen in der Form von Rundungsfehlern vor.

2.3 Chaotische Dynamik

Durch die Existenz von Rauschen in Zeitreihen geht Information über die Dynamik des Systems verloren. Jedoch kann auch bei rein deterministischen Systemen durch Nichtlinearitäten ein Informationsverlust auftreten. Dieses Phänomen führt zu dem Begriff des deterministischen Chaos. Es gibt keine einheitliche Definition von Chaos, jedoch wird chaotisches Verhalten am häufigsten durch die sensitive Abhängigkeit von den gegebenen Anfangsbedingungen charakterisiert [2]. Das bedeutet, daß zwei beliebig dicht im Phasenraum benachbarte Trajektorien in

ihrer zeitlichen Entwicklung voneinander getrennt werden können. Ein derartiges Verhalten kann durch positive *Lyapunov-Exponenten* λ_i^+ quantifiziert werden.

Der Lyapunov- (oder charakteristische) Exponent gibt die durchschnittliche exponentielle Wachstumsrate an, mit der sich eine kleine Störung im Zustandsraum zeitlich entwickelt [3]. Der multiplikative Ergodensatz von Oseledec [4] liefert die mathematische Herleitung dieser Größe.

Dazu betrachtet man eine meßbare Abbildung f auf einem Unterraum M des Zustandsraums (hier \mathbb{R}^d). Desweiteren sei T eine lineare Abbildung auf demselben Raum. Diese ist in der Regel durch die Differentialmatrix der Flußabbildung $f^t(\mathbf{x})$ gegeben:

$$T_{\mathbf{x}}^t = D_{\mathbf{x}}f^t = \left(\frac{\partial f_i^t(\mathbf{x})}{\partial x_j} \right)_{i,j}. \quad (2.4)$$

Die linearisierte Abbildung $T_{\mathbf{x}}^t$ beschreibt damit die zeitliche Entwicklung eines Tangentialvektors. Die Lyapunov-Exponenten λ_i sind als die Logarithmen der Eigenwerte der Matrix

$$\Lambda_{\mathbf{x}} = \lim_{t \rightarrow \infty} [T_{\mathbf{x}}^{t*} T_{\mathbf{x}}^t]^{\frac{1}{2t}}$$

definiert. Sie können damit, unter Verwendung von (2.4), als die logarithmischen Wachstumsraten

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t} \log \|D_{\mathbf{x}}f^t \mathbf{u}\|, \quad \mathbf{u} \in E_{\mathbf{x}}^{(i)} \setminus E_{\mathbf{x}}^{(i+1)} \quad (2.5)$$

angesehen werden, wobei $E_{\mathbf{x}}^{(i)}$ der Unterraum des \mathbb{R}^d ist, der zu den Eigenwerten $\leq \exp(\lambda_i)$ korrespondiert [3, 5].

2.4 Entropie und Information

Um Informationsverluste zu analysieren, ist es nötig, ein Maß für die Information zu besitzen. Fast alle derartigen Maße basieren auf Shannons Definition der Entropie [6]

$$H(Y) = - \sum_{\alpha=1}^N P(Y = y_{\alpha}) \log P(Y = y_{\alpha}), \quad (2.6)$$

wobei $P(Y = y_{\alpha})$ die Wahrscheinlichkeit ist, daß die Zufallsvariable Y einen der N verschiedenen Werte y_{α} annimmt ¹. Damit stellt die Entropie $H(Y)$ die Menge an

¹In der weiteren Arbeit werden auch die Kurzformen $P(y_{\alpha}) = P(Y = y_{\alpha})$ im diskreten und $P(A) = P(Y \in A)$ im kontinuierlichen Fall verwendet.

Information dar, die zusätzlich notwendig ist, um die Zufallsfunktion Y eindeutig zu beschreiben.

Zusätzliche Information kann durch die Kenntnis einer weiteren Zufallsvariablen X gegeben sein. Die Menge an Unkenntnis über Y , die trotz des Wissens über X besteht, wird durch die *bedingte Entropie*

$$H(Y|X) = - \sum_{\alpha=1}^N \sum_{\beta=1}^M P(Y = y_{\alpha}, X = x_{\beta}) \log P(Y = y_{\alpha}|X = x_{\beta}) \quad (2.7)$$

angegeben. Hierbei ist $P(Y = y_{\alpha}|X = x_{\beta})$ die bedingte Wahrscheinlichkeit von y_{α} bei gegebenem x_{β} . Die bedingte Entropie hat zwei wichtige Eigenschaften [7]:

- Ist in X Information über Y enthalten, so ist die bedingte Entropie $H(Y|X)$ kleiner als die Einzelentropie $H(Y)$:

$$H(Y|X) \leq H(Y). \quad (2.8)$$

Die Gleichheit besteht im Falle der statistischen Unabhängigkeit von Y bzgl. X .

- Kann Y als Funktion von X dargestellt werden, so verschwindet die bedingte Entropie:

$$H(Y|X) = 0, \quad \text{falls } Y = f(X). \quad (2.9)$$

Durch diese beiden Eigenschaften ist die bedingte Entropie eine geeignete Größe, um Zusammenhänge und Informationsflüsse zwischen verschiedenen Zeitreihen zu quantifizieren.

Die Transinformation

$$I(X, Y) = \sum_{\alpha=1}^N \sum_{\beta=1}^M P(Y = y_{\alpha}, X = x_{\beta}) \log \frac{P(Y = y_{\alpha}, X = x_{\beta})}{P(Y = y_{\alpha})P(X = x_{\beta})} \quad (2.10)$$

gibt schließlich die Menge an Information an, die beiden Größen X und Y gemeinsam ist. Sie kann mittels

$$I(X, Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$$

durch Entropie und bedingte Entropie ausgedrückt werden. Sind X und Y voneinander unabhängig, so verschwindet die Transinformation.

2.5 Kolmogorov-Sinai-Entropie

Der informationstheoretische Ansatz setzt in der obigen Form eine endliche oder zumindest abzählbare Zustandsmenge voraus. Ein Problem besteht nun darin, eine Entropie bezüglich eines dynamischen Systems mit kontinuierlichem Zustandsraum zu definieren. Den Ideen von Kolmogorov (1958) und Sinai (1959) folgend, wird dazu der Zustandsraum, bzw. der Teilraum M , in einzelne Unterräume aufgeteilt [8].

Eine derartige *Partitionierung* $\mathfrak{M} = \{M_1, \dots, M_k\}$ des Wahrscheinlichkeitsraums (M, \mathcal{A}, P) ist eine Sammlung von Mengen $M_i \in \mathcal{A}$, deren Vereinigung M ist. Aus diesen Mengen kann eine Sub- σ -Algebra $\mathcal{B}(\mathfrak{M})$ gewonnen werden [9], so daß wiederum ein Wahrscheinlichkeitsmaß definierbar ist. Eine *Verfeinerung* $\mathfrak{N} = \{N_1, \dots, N_l\}$ der Partition \mathfrak{M} ist dann gegeben, wenn jedes Element aus \mathfrak{M} eine Vereinigung von Mengen aus \mathfrak{N} ist. Die gemeinsame Partition von zwei endlichen Partitionen \mathfrak{M} und \mathfrak{N} ist

$$\mathfrak{M} \vee \mathfrak{N} = \{M_i \cap N_j : 1 \leq i \leq k, 1 \leq j \leq l\}.$$

Jede Teilmenge einer Partition \mathfrak{M} besitzt eine Wahrscheinlichkeit $P(M_i)$. Folglich kann eine Entropie der Partition durch

$$H(\mathfrak{M}) = - \sum_{i=1}^k P(M_i) \log P(M_i) \quad (2.11)$$

definiert werden. Ist der Zustandsraum durch ein dynamisches System $\{f^t\}_{t \in \mathbb{N}_0}$ bestimmt und ist \mathfrak{M} eine endliche Partition dieses Zustandsraums, so ist durch

$$\mathfrak{M}^{(n)} = \mathfrak{M} \vee f^{-1} \mathfrak{M} \vee \dots \vee f^{-n+1} \mathfrak{M} \quad (2.12)$$

eine Verfeinerung von \mathfrak{M} gegeben. Durch die Grenzwerte

$$h(P, \mathfrak{M}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathfrak{M}^{(n)}) = \lim_{n \rightarrow \infty} [H(\mathfrak{M}^{(n+1)}) - H(\mathfrak{M}^{(n)})] \quad (2.13)$$

und

$$h(P) = \lim_{\epsilon \rightarrow 0} h(P, \mathfrak{M}), \quad (2.14)$$

wobei ϵ in diesem Fall den größten auftretenden Durchmesser der Mengen M_i bezeichnet, ist die mittlere Verlustrate an Information bzgl. des Wahrscheinlichkeitsmaßes P und des dynamischen Systems gegeben. Diese Rate heißt *Kolmogorov-Sinai-Entropie* [3]. Sie ist eng mit dem Lyapunov-Spektrum verknüpft; ist P ein

SRB-Maß (Anhang B), so entspricht die Kolmogorov-Sinai-Entropie der Summe der positiven Lyapunov-Exponenten [10]:

$$h(P) = \sum \lambda_i^+. \quad (2.15)$$

Diese Gleichung gilt in den meisten Fällen auch für physikalische Maße [3].

Die Gleichheit (2.15) kann anschaulich wie folgt hergeleitet werden:

Ein dynamisches System kann mit einer gewissen (Meß-)Genauigkeit ϵ beschrieben werden. Dementsprechend wird der Zustandsraum in d -dimensionalen Würfeln W_ϵ der Kantenlänge ϵ aufgeteilt.

Die Volumenänderung unter der differenzierbaren Abbildung f^t ist über den Transformationssatz für Lebesguemaße durch die Jacobi-Determinante

$$\|J_{\mathbf{x}}^t\| = \left| \det \left(\frac{\partial f_i^t(\mathbf{x})}{\partial x_j} \right) \right|$$

und damit durch das Volumen

$$V(f^t(M)) = \mu_L^d(f^t(M)) = \int_M \|J_{\mathbf{x}}^t\| (d\mu_L^d)(\mathbf{x})$$

gegeben [11], wobei μ_L^d das d -dimensionale Lebesguemaß ist.

Das Bild eines ϵ -Würfels erstreckt sich folglich über ein Volumen von ca. $\epsilon^d \|J_{\mathbf{x}}^t\|$. In den expandierenden, instabilen Richtungen, also jenen mit positivem lokalen charakteristischen Exponenten $\lambda_{\mathbf{x}}^+$, wird ein Würfel um den Faktor $e^{\lambda_{\mathbf{x}}^+}$ gestreckt. Damit werden bei gleicher Meßgenauigkeit ϵ dementsprechend mehr ϵ -Würfel benötigt, um das Bild des Würfels zu überdecken; es geht Information verloren. In den anderen Richtungen kann das Bild weiterhin mit einer Genauigkeit von ϵ lokalisiert werden. Der über den ganzen Zustandsraum gemittelte Informationsverlust ist damit bei verschwindendem ϵ durch $\int \sum_i \lambda_{\mathbf{x},i}^+ dP$ gegeben. In einem ergodischen System kann dieser über den Phasenraum gemittelte Erwartungswert durch ein Zeitmittel über die lokalen Informationsverluste entlang einer Trajektorie ersetzt werden. Dies führt zur Definition der Lyapunov-Exponenten (2.5).

3 Informationsverluste über einzelne Meßgrößen

In diesem Kapitel wird der Einfluß nichtlinearer Dynamik auf die Information bzgl. einer festen Beobachtungsgröße behandelt. Dies geschieht mittels Übertragung der im vorherigen Abschnitt vorgestellten Ideen und Ergebnisse bzgl. der Kolmogorov-Sinai-Entropie. Hier wird nun untersucht, wie trotz der Kenntnis eines d -dimensionalen Systemzustands Information über die zeitliche Entwicklung *einer* Meßgröße verloren gehen kann. Insbesondere wird dabei der Fall fest vorgegebener Meßrichtungen betrachtet. Diese liegen dann vor, wenn die Systemzustände nur durch die Messung festgelegter Größen bestimmt werden können. In diesem Kapitel wird eine auf diesen Fall angepaßte Entropie eingeführt, die es erlaubt, den Informationsverlust zu quantifizieren [12].

3.1 Der eindimensionale Fall

Im Fall $d = 1$ ist das dynamische System durch eine Meßgröße bestimmt und kann somit durch eine Abbildung $f^t : \mathbb{R} \rightarrow \mathbb{R}$ beschrieben werden. Daher ist der Informationsverlust bzgl. dieser einen Größe derselbe wie jener bzgl. des gesamten Systems. In diesem Fall entspricht die Problemstellung der im letzten Kapitel beschriebenen Situation zur Herleitung von Lyapunov-Exponenten und Kolmogorov-Sinai-Entropie. Dieser Abschnitt dient vor allem dazu, die entsprechenden Ideen, welche auch für den mehrdimensionalen Fall von Bedeutung sind, zu veranschaulichen.

Dazu sei f eine nichtlineare Funktion, die ein Intervall I auf sich selbst abbildet. Diese kann im Fall nichtlinearer dynamischer Systeme als Funktion $f = f^1$ aus (2.2) angenommen werden. Zur Bestimmung des Informationsverlustes wird untersucht, wie Teil-Intervalle $I_\delta(x) = [x - \delta, x + \delta]$ abgebildet werden. In Abbildung 1 ist ein willkürliches Beispiel für eine derartige Funktion zu sehen.

Für kleine δ kann eine Linearisierung der Funktion f vorgenommen werden. Die Größe von $f(I_\delta(x))$ ist dann durch $2\epsilon \simeq 2\delta|f'(x)|$ gegeben. Ist $|f'(x)| > 1$, wird das Intervall folglich gestreckt, im Fall $|f'(x)| < 1$ gestaucht. Der Lyapunov-Exponent ergibt sich als Erwartungswert $\lambda = E(\log |f'(x)|)$.

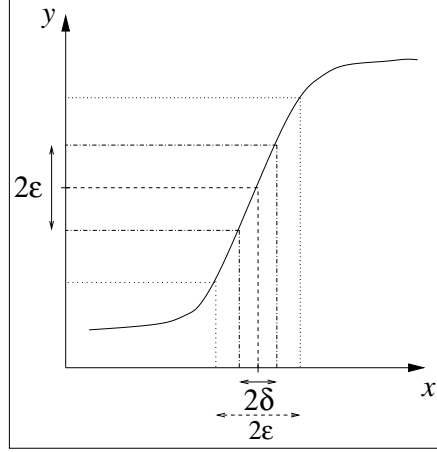


Abbildung 1: Intervalle werden durch eine nichtlineare Funktion gestreckt.

Für die Bestimmung der Kolmogorov-Sinai-Entropie wird I in ϵ -Intervalle aufgeteilt, so daß man eine Partition \mathfrak{M}_ϵ erhält. Die Urbilder dieser ϵ -Intervalle $I_\epsilon(f(x))$ werden untersucht, in diesem Fall also $I_\delta(x)$. Durch $f^{-1}(\mathfrak{M}_\epsilon)$ ist eine neue Partition gegeben. Die Entropie der gemeinsamen Partition ist

$$H(\mathfrak{M}_\epsilon \vee f^{-1}(\mathfrak{M}_\epsilon)) = H(\mathfrak{M}_\epsilon) + H(f^{-1}(\mathfrak{M}_\epsilon) | \mathfrak{M}_\epsilon). \quad (3.1)$$

Zur Bestimmung der Kolmogorov-Sinai-Entropie werden nach (2.13) Entropiedifferenzen zu aufeinanderfolgenden Zeitpunkten berechnet. Die Differenz

$$H(\mathfrak{M}_\epsilon^{(1)}) - H(\mathfrak{M}_\epsilon^{(0)}) = H(f^{-1}(\mathfrak{M}_\epsilon) | \mathfrak{M}_\epsilon)$$

bezüglich des ersten Zeitschritts ist mit den Bezeichnungen von Gleichung (2.12) eine bedingte Entropie. Diese berechnet sich gemäß Gleichung (2.7) als negativer Erwartungswert der logarithmischen bedingten Wahrscheinlichkeiten

$$H(f^{-1}(\mathfrak{M}_\epsilon) | \mathfrak{M}_\epsilon) = - \sum_{i,j} P(I_{\epsilon,i} \cap I_{\delta,j}) \log \frac{P(I_{\epsilon,i} \cap I_{\delta,j})}{P(I_{\epsilon,i})}.$$

Das Wahrscheinlichkeitsmaß P kann durch das auf 1 normierte Lebesguemaß ausgedrückt werden. Dann nimmt die bedingte Wahrscheinlichkeit, wie oben gezeigt, im Fall $|f'(x)| \leq 1$ den Wert 1 an, andernfalls $|f'(x)|^{-1}$.

Die Grenzwertbildung bei unendlich wiederholter Iteration von f^{-1} und gegebenenfalls $\epsilon \rightarrow 0$ führt dann zur Kolmogorov-Sinai-Entropie. Diese entspricht dem positiven Lyapunov-Exponenten; ist der Lyapunov-Exponent jedoch negativ, so verschwindet die Kolmogorov-Sinai-Entropie.

Für die weitere Argumentation sind zwei Punkte von Bedeutung:

- Informationsverluste werden durch den Erwartungswert negativer Logarithmen von bedingten Wahrscheinlichkeiten, also bedingte Entropien, beschrieben.
- Die bedingten Wahrscheinlichkeiten können mittels einer Linearisierung der Funktion f berechnet werden.

3.2 Der mehrdimensionale Fall

Im d -dimensionalen Fall ($d \geq 1$) wird nun der Informationsverlust bei Funktionen $f_j : \mathbb{R}^d \rightarrow I \in \mathbb{R}$ untersucht. Dies ist ein anderer Ansatz als bei der Kolmogorov-Sinai-Entropie, bei der dynamische Systeme (also Funktionen $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$) betrachtet werden. Die Herleitung erfolgt jedoch weitgehend analog zu der im eindimensionalen Fall vorgestellten Methode.

Zunächst wird die Funktion f_j linearisiert. Dies erfolgt nicht mehr mittels der Differentialmatrix, sondern durch das totale Differential

$$df_j(\mathbf{x}) = \sum_{i=1}^d \frac{\partial f_j(\mathbf{x})}{\partial x_i} dx_i.$$

Der konstante Term der Linearisierung ist für den Informationsverlust bedeutungslos. Daher wird die Abbildung g_j , welche der j -ten Komponente der linearisierten Flußabbildung entspricht,

$$g_j(\mathbf{x}) = \sum_{i=1}^d a_i x_i \quad \text{mit} \quad a_i = a_{i,j} := \frac{\partial f_j(\mathbf{x})}{\partial x_i}$$

gesetzt.

Im zweiten Schritt wird dann untersucht, welcher Teil einer ϵ -Umgebung $U_\epsilon(\mathbf{x})$ durch die linearisierte Abbildung auf ein ϵ -Intervall $I_\epsilon(f_j(\mathbf{x}))$ abgebildet wird. Dabei reicht es aus, eine Umgebung des Ursprungs zu betrachten, und $I_\epsilon(f_j(\mathbf{x})) = [-\epsilon, \epsilon]$ zu wählen. Durch $g_j(\mathbf{x}) = \pm\epsilon$ werden zwei $(d-1)$ -dimensionale Hyperebenen beschrieben.

Für den zweidimensionalen Fall sind dies zwei Geraden (Abbildung 2); als ϵ -Umgebung wurde dort ein Kreis mit Radius ϵ gewählt, was einer Verwendung der

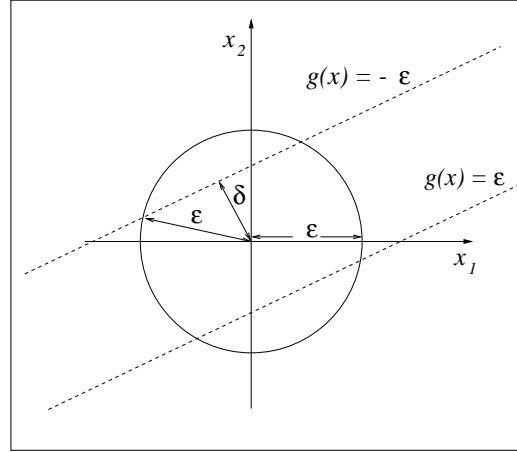


Abbildung 2: Der Teil der ϵ -Kugel, welcher auf I_ϵ abgebildet wird, befindet sich zwischen den beiden Geraden $g_j(\mathbf{x}) = \pm\epsilon$.

euklidischen Abstandsnorm entspricht. Alle Punkte der Kreisfläche, oder im allgemeineren Fall der d -dimensionalen Umgebung, die zwischen diesen Hyperebenen liegen, werden folglich auf I_ϵ abgebildet. Der Abstand 2δ der beiden Hyperebenen ist gegeben durch²

$$\delta = \epsilon \cdot \left(\sqrt{\sum_{i=1}^d a_i^2} \right)^{-1}. \quad (3.2)$$

Das bedeutet, je größer die Beträge der partiellen Ableitungen a_i werden, desto kleiner wird δ . Im eindimensionalen Fall wird diese Gleichung zu

$$\delta = \frac{\epsilon}{|f'(x)|},$$

was im vorangegangenen Abschnitt bereits verwendet wurde.

Danach muß die bedingte Wahrscheinlichkeit berechnet werden. Diese ist der Quotient aus der Wahrscheinlichkeit des Umgebungsteils U_δ zwischen den gegebenen Hyperebenen und der Wahrscheinlichkeit der Gesamtumgebung U_ϵ

$$P(I_\epsilon|U_\epsilon) = \frac{P(U_\delta)}{P(U_\epsilon)}. \quad (3.3)$$

Mittels der auf diese Weise definierten bedingten Wahrscheinlichkeit kann, wie-

²Dies kann mittels Lagrange-Multiplikatoren aus den Koeffizienten a_i von g_j berechnet werden. Dazu wird der Abstand aller Punkte zum Ursprung mit der Nebenbedingung $g_j(\mathbf{x}) = \epsilon$ minimiert.

derum analog zum eindimensionalen Fall, eine Entropie definiert werden

$$H(f_j, \epsilon) = -E \left(\log \frac{P(U_\delta)}{P(U_\epsilon)} \right). \quad (3.4)$$

Abschließend wird der Grenzwert für verschwindende Umgebungsgrößen ϵ gebildet

$$h^*(f_j) = \lim_{\epsilon \rightarrow 0} H(f_j, \epsilon). \quad (3.5)$$

Im ergodischen Fall kann die Erwartungswertbildung in (3.4) als zeitliches Mittel entlang einer Trajektorie durchgeführt werden.

3.3 Abschätzung der bedingten Wahrscheinlichkeit und Wahl der ϵ -Umgebung

Für die Berechnung von $h^*(f_j)$ ist gemäß Gleichung (3.4) notwendig, bedingte Wahrscheinlichkeiten zu bestimmen. Dazu müssen nach (3.3) die Wahrscheinlichkeiten der ϵ -Umgebung und die der Teilmenge, die auf I_ϵ abgebildet wird, berechnet werden. Deren Quotient ergibt sich unter Verwendung der durch $\{f^t\}$ bestimmten, invarianten Dichte ρ durch

$$\eta_d = \frac{P(U_\delta)}{P(U_\epsilon)} = \frac{\int_{U_\delta} \rho(\mathbf{x}) d\mathbf{x}}{\int_{U_\epsilon} \rho(\mathbf{x}) d\mathbf{x}}. \quad (3.6)$$

Dieser Quotient hängt von der Wahl der Umgebung U_ϵ ab, welche bisher noch nicht genauer spezifiziert worden ist. Mit der Annahme, daß für kleine ϵ die Dichte $\rho(\mathbf{x})$ innerhalb der ϵ -Umgebung als konstant angenommen und somit P wieder durch das normierte Lebesguemaß ausgedrückt werden kann, ist die bedingte Wahrscheinlichkeit durch den Quotienten der Volumina $V(U_\delta)/V(U_\epsilon)$ gegeben.

Während $\eta_d(U_\epsilon)$ in der Regel numerisch berechnet werden muß, ist es möglich, wenn als ϵ -Umgebung eine d -dimensionale Kugel B_ϵ mit Radius ϵ (wie in Abbildung 2) gewählt wird, $\eta_d(B_\epsilon)$ analytisch zu bestimmen. In diesem Fall hängt der Quotient (3.6) nur von dem Abstand der gegebenen Hyperebenen $g(\mathbf{x})$ und nicht von deren Lage ab.

Berechnung von $\eta_d(B_\epsilon)$

Das Volumen einer d -dimensionalen Kugel mit Radius R lässt sich durch

$$V_{B_R}(d) = \beta(d) \cdot R^d$$

ausdrücken, wobei $\beta(d)$ das Verhältnis des Volumens der Kugel zu dem des einschließenden d -dimensionalen Würfel angibt und von R unabhängig ist. Unter Verwendung des Satzes von Fubini [11] ist das Kugelvolumen durch die Aufintegration $(d-1)$ -dimensionaler Kugeln gegeben

$$V_{B_R}(d) = 2\beta(d-1) \int_0^R \sqrt{(R^2 - r^2)^{d-1}} \, dr. \quad (3.7)$$

Die Berechnung eines durch zwei parallele Hyperebenen beschränkten Teilvolumens erfolgt analog, indem nur die Integrationsgrenzen verändert werden. Damit lässt sich die bedingte Wahrscheinlichkeit (3.6) durch

$$\eta_d(B_\epsilon) = \frac{P(U_\delta)}{P(B_\epsilon)} = \frac{\int_0^\delta (\sqrt{\epsilon^2 - r^2})^{d-1} \, dr}{\int_0^\epsilon (\sqrt{\epsilon^2 - r^2})^{d-1} \, dr} \quad (3.8)$$

ausdrücken. Die Integrale lassen sich nun analytisch berechnen. Mit dem Binomialkoeffizienten $c(p, q) = \binom{p}{q}$ und den Abkürzungen $m = \frac{d-1}{2}$,

$$b(m, i) = \frac{c(2m+1, m)}{2i \cdot 4^{m-i+1} \cdot c(2i-1, i-1)} \quad \text{und}$$

$$S(R, x) = \sqrt{R^2 - x^2}$$

ist das bestimmte Integral:

$$\int_0^x S(R, r)^d \, dr = \begin{cases} \sum_{i=0}^{d/2} c(\frac{1}{2}d, i) \cdot \frac{-1^i}{2i+1} \cdot R^{2i+1} \cdot x^{2(d/2-i)} & : \text{ falls } d \text{ gerade} \\ b(m, 1) \arcsin\left(\frac{x}{R}\right) + \sum_{i=1}^{m+1} b(m, i) S(R, x)^{2i-1} x & : \text{ falls } d \text{ ungerade.} \end{cases}$$

Der Quotient η_d hängt von dem Verhältnis von δ zu ϵ , sowie von der Dimension d ab. Dabei ist δ durch ϵ und die Koeffizienten a_i der linearen Abbildung g bestimmt. Für verschiedene Dimensionen d wurde die Funktion $\eta_d = \eta_d(\delta/\epsilon)$

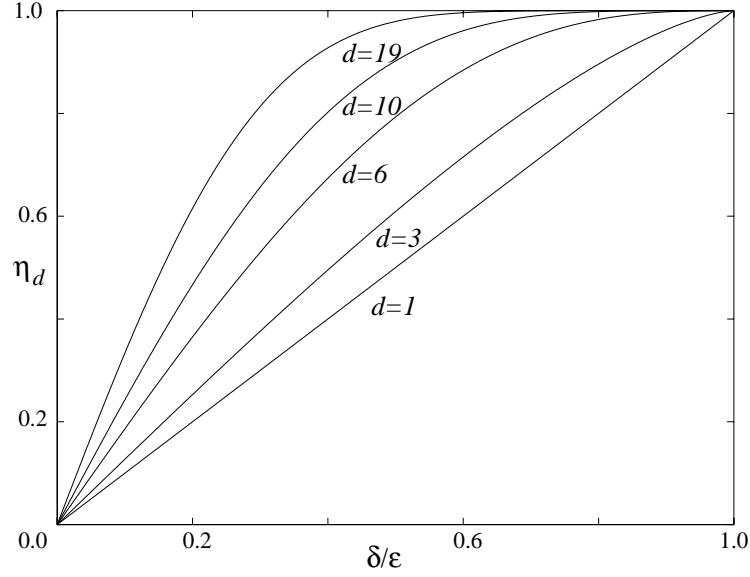


Abbildung 3: η_d als Funktion von $\frac{\delta}{\epsilon}$ bei verschiedenen Dimensionen d .

mittels obiger Gleichung berechnet. In Abbildung 3 ist η_d für $d=1, 3, 6, 10$ und 19 als Funktion dieses Quotienten aufgetragen.

Im eindimensionalen Fall ist $\eta_1\left(\frac{\delta}{\epsilon}\right) = \frac{\delta}{\epsilon}$. Da dabei für verschwindende Intervallgrößen $\epsilon = \delta|f'|$ gilt, ist dies wiederum das bereits bekannte Ergebnis für die bedingten Wahrscheinlichkeiten.

Abhängigkeit von den Koordinatenachsen

Mit wachsender Dimension wird η_d jedoch für jedes feste $\frac{\delta}{\epsilon}$ größer. Ist nun ein Koeffizient der linearen Abbildung $a_i^* = 0$, so hat er keinen Einfluß auf δ . Außerdem hat die zugehörige Komponente x_i^* keinen Einfluß auf y . Dennoch würde die bedingte Wahrscheinlichkeit aufgrund der Dimensionsabhängigkeit von η_d anwachsen. Da diese nach ihrer Definition ein Maß für den lokalen Informationsverlust ist, könnte demnach durch Hinzunahme einer „Nullinformation“ Wissen vernichtet werden. Dies steht jedoch im Widerspruch zur Definition von Information.

Der Grund für diesen scheinbaren Widerspruch liegt in der Wahl der ϵ -Kugel als Umgebung. Damit wird implizit angenommen, daß der Informationsverlust von der Wahl der Koordinatenachsen unabhängig ist. Dies ist aber nicht der Fall, wenn die Meßgenauigkeit ϵ nur in Richtung der einzelnen Koordinatenachsen

festgelegt wird. Da die Koeffizienten a_i von g durch die gewählten Koordinatenachsen festgelegt sind, ist der Informationsverlust bzgl. einer Meßgröße dann von der Orientierung des Koordinatensystems abhängig.

In diesem Fall müssen die ϵ -Umgebungen daher d -dimensionale Würfel sein, deren Kanten parallel zu den Basisvektoren des Koordinatensystems liegen. Dann hängt η_d wieder von der Lage der Hyperebenen ab. Diese würden im Fall $a_i^* = 0$ parallel zur Koordinatenachse \mathbf{e}_j liegen, und somit keinen Einfluß auf die bedingte Wahrscheinlichkeit haben. Da der Beitrag der Komponente x_i^* auf die Volumina von U_ϵ und U_δ jeweils aus dem Faktor 2ϵ besteht, kann er in Gleichung (3.6) herausgekürzt werden.

Aufgrund der Abhängigkeit des Informationsverlustes von den Koordinatenachsen werden diese während der gesamten zeitlichen Entwicklung als fest angenommen. Dies entspricht der Vorstellung, daß das dynamische System nur durch Messung fester Größen beobachtet werden kann. Die Verwendung d -dimensionaler Kugeln als Umgebung entspricht einer Mittelung über alle möglichen Orientierungen des Koordinatensystems.

3.4 Zusammenhang mit der Kolmogorov-Sinai-Entropie

Die oben eingeführte Entropie $h^*(f_j)$ gibt an, wieviel Information über *eine* Meßgröße trotz Kenntnis eines kompletten Systemzustands im Durchschnitt durch die zeitliche Entwicklung des Systems verloren geht. Dies ist ein Unterschied zur Kolmogorov-Sinai-Entropie $h(P)$, durch die der Informationsverlust bzgl. des ganzen Systems angegeben wird. Die folgenden Überlegungen sollen Anhaltspunkte für einen Vergleich zwischen beiden Entropien liefern.

Über eine Komponente des Systems wird in der Regel nicht mehr Information verloren als über das gesamte System. Numerische Untersuchungen zeigen, daß in mehrdimensionalen Systemen in den meisten Fällen $h^*(f_j)$ für jede feste Koordinatenachse des Systems kleiner als die Kolmogorov-Sinai-Entropie ist,

$$h^*(f_j) \leq h(P).$$

Nun kann weiterhin untersucht werden, wie groß der Informationsverlust über das gesamte dynamische System bei unabhängigen Messungen ist. Das bedeutet: die zeitliche Entwicklung des Zustandsvektors ist nicht mehr bekannt, sondern

nur noch die der einzelnen Komponenten, also der Projektionen auf die einzelnen Meßachsen. Dies ist beispielsweise dann von Interesse, wenn die einzelnen Meßgrößen nicht gleichzeitig gemessen werden können.

In diesem Fall ist der Informationsverlust durch die Summe der Entropie $h^*(f_j)$, $j = 1, \dots, d$ bzgl. jeder der d Meßrichtungen gegeben. Ein Zustand, der anfangs mit einer Genauigkeit von 2ϵ in jeder Richtung bestimmt werden kann, ist nach der durch die Dynamik bestimmten zeitlichen Entwicklung nur noch mit einer Genauigkeit von $\max(2\epsilon, 2\epsilon \sum |\partial f_j / \partial x_i|)$ in den entsprechenden Richtungen lokalisierbar. Dies entspricht einer Abbildung auf einen d -dimensionalen Quader Q mit einer minimalen Kantenlänge vom 2ϵ . Das Bild eines ϵ -Würfels $C_\epsilon(\mathbf{x})$ wird für jedes \mathbf{x} von dem entsprechenden Quader $Q(f(\mathbf{x}))$ eingeschlossen,

$$f(C_\epsilon(\mathbf{x})) \subset Q(f(\mathbf{x})).$$

An verschiedenen Beispielen durchgeführte numerische Untersuchungen deuten darauf hin, daß die Summe über alle $h^*(f_j)$ in der Regel nicht kleiner als die Kolmogorov-Sinai-Entropie ist, also:

$$h(P) \leq \sum_{j=1}^d h^*(f_j). \quad (3.9)$$

4 Bestimmung der bedingten Entropie aus Zeitreihen

In diesem Kapitel wird eine modifizierte bedingte Entropie eingeführt und gezeigt, wie diese aus Zeitreihen bestimmt werden kann [12]. Folgende Fragen sollen damit beantwortet werden:

- Ist in einer Zeitreihe Information über eine andere enthalten?
- Reichen die vorhandenen Meßgrößen aus, um die zeitliche Entwicklung einer Observablen vorherzusagen?
- Welche Meßgrößen sind dazu am besten geeignet?

4.1 Abschätzung bedingter Entropien

In diesem Abschnitt wird ein Verfahren vorgestellt, mittels dessen bedingte Wahrscheinlichkeiten und damit auch bedingte Entropien aus Zeitreihen abgeschätzt werden können. Dabei werden keine Annahmen über ein zugrundeliegendes dynamisches System gemacht.

Es werden daher zunächst Zufallsvariablen X_t und Y_t betrachtet, die die Zeitreihen $\{\mathbf{x}_t\}_{t \leq T}$, bzw. $\{y_t\}_{t \leq T}$ erzeugen. Dabei nimmt t nur diskrete Werte an, welche ein Vielfaches einer Abtastzeit t_s sind. Durch $N_T = T/t_s$ ist die Länge der Zeitreihe gegeben. Die Realisierungen \mathbf{x}_t stammen aus dem \mathbb{R}^d , die y_t aus \mathbb{R} . Für die Berechnungen ist unerheblich, ob Y eine tatsächlich neue Meßgröße beschreibt oder aber eine zeitversetzte Komponente darstellt, so daß z.B. $y_t = \mathbf{x}_{j,t+\Delta\tau}$ ist. Weiterhin wird angenommen, daß die Zeitreihen in allen Komponenten normiert sind (in der Regel auf das Intervall $[0, 1]$).

Da die Wahrscheinlichkeit eines einzelnen Punktes aus \mathbb{R}^d in der Regel verschwindet, werden wiederum ϵ -Umgebungen der Punkte gebildet. Deren Wahrscheinlichkeit ist dann durch $P(X \in U_\epsilon(\mathbf{x}))$ gegeben. Diese kann aus einer Zeitreihe einfach abgeschätzt werden, indem sie mit der relativen Häufigkeit, mit der die Punkte der Zeitreihe innerhalb dieser Umgebung liegen, identifiziert wird:

$$P_\epsilon(\mathbf{x}) = \frac{1}{N_T} \#\{s : \|\mathbf{x}_s - \mathbf{x}\| \leq \epsilon, 0 \leq s \leq T\}.$$

Es werden also die Punkte gezählt, deren Abstand von \mathbf{x} kleiner als ϵ ist, und deren Anzahl durch die Gesamtzahl an Punkten in der vorhandenen Zeitreihe geteilt. Die Abstandsbestimmung erfolgt gemäß der in Abschnitt 3.3 vorgestellten Überlegungen mittels der Maximumsnorm. Die Grenzwertbildung

$$\rho(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \frac{\mathbf{1}_{U_\epsilon(\mathbf{x})}(\mathbf{x}_t)}{\mu_L^d(U_\epsilon(\mathbf{x}))} \quad (4.1)$$

führt dann zu der Wahrscheinlichkeitsdichte ρ bezüglich des d -dimensionalen Lebesguemaßes, welches das Volumen der Umgebungen $U_\epsilon(\mathbf{x})$ angibt. Dabei ist $\mathbf{1}_A$ die Indikatorfunktion bzgl. einer Menge A , d.h. $\mathbf{1}_A(\mathbf{x}) = 1$ wenn $\mathbf{x} \in A$ und sonst 0.

Die Bestimmung der bedingten Wahrscheinlichkeit $P(y|\mathbf{x})$ erfolgt analog. Es werden also die Punktepaare $(\mathbf{x}_{s'}, y_{s'})$ gezählt, deren Abstand von (\mathbf{x}, y) kleiner als ϵ ist. Durch die Verwendung der Maximumsnorm ist dies die Anzahl der $(d+1)$ -dimensionalen Punkte, die in einem Würfel mit Kantenlänge ϵ liegen (Abbildung 4). Die Division durch die Gesamtzahl vorhandener Punkte ergibt eine gemeinsame Wahrscheinlichkeit

$$P_\epsilon(\mathbf{x}, y) = \frac{1}{N_T} \#\{s' : \|\mathbf{x}_{s'} - \mathbf{x}\| \leq \epsilon \wedge |y_{s'} - y| \leq \epsilon, 0 \leq s' \leq T\}.$$

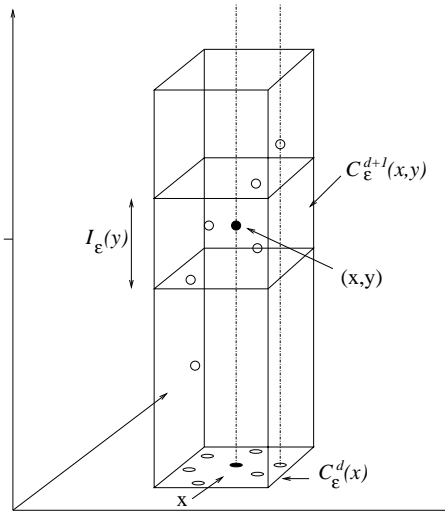


Abbildung 4: Abschätzung der bedingten Wahrscheinlichkeit $P_\epsilon(y|\mathbf{x})$. Sie erfolgt durch Division der Anzahl der Punkte in $C_\epsilon^{d+1}(\mathbf{x}, y)$ durch die Anzahl der Punkte in $C_\epsilon^d(\mathbf{x})$.

Der Quotient

$$P_\epsilon(y|\mathbf{x}) = \frac{P_\epsilon(\mathbf{x}, y)}{P_\epsilon(\mathbf{x})} \quad (4.2)$$

ist dann eine Abschätzung für die bedingte Wahrscheinlichkeit. Die Grenzwertbildung für unendliche Zeitreihen und verschwindende Umgebungsgrößen ergibt dann eine bedingte Dichte

$$\rho(y|\mathbf{x}) = \frac{\rho(\mathbf{x}, y)}{\rho(\mathbf{x})}. \quad (4.3)$$

Diese wurde von Shannon verwendet, um die Definition der bedingten Entropie auf den kontinuierlichen Fall übertragen zu können [6, 7]. Diese bedingte Entropie ist dann gegeben durch

$$h(Y|X) = - \int \log \rho(y|\mathbf{x}) \rho(\mathbf{x}, y) d\mathbf{x} dy. \quad (4.4)$$

4.2 Korrektur der nichtlinearen Einflüsse

Aus den Untersuchungen des vorangegangenen Kapitels ergibt sich, daß die Entropie (4.4) nicht mehr die Eigenschaft (2.9) besitzt; $h(Y|X)$ muß nicht notwendigerweise verschwinden, wenn Y eine Funktion von X ist. Es wurde gezeigt, daß bei analoger Bildung von ϵ -Umgebungen bei nichtlinearen Zusammenhängen Information über eine Meßgröße verloren gehen kann. Geht also Y durch eine nichtlineare Funktion f aus X hervor, so ist

$$h(Y|X) = h^*(f). \quad (4.5)$$

Dieser Wert kann aber positive Werte annehmen. Somit kann aus $h(Y|X) \neq 0$ nicht geschlossen werden, daß Y keine eindeutige Funktion von X ist, da Einflüsse von Nichtlinearitäten und die damit verbundenen Informationsverluste nicht berücksichtigt werden.

Wenn nun aber aus einer Zeitreihe auf Vorhersagbarkeit im Sinne einer Modellierbarkeit durch eine stetige Funktion geschlossen werden soll, ist Eigenschaft (2.9) von zentraler Bedeutung. Man könnte daher eine bedingte Entropie durch die Differenz $h(Y|X) - h^*(f)$ definieren. Allerdings kann bei der Analyse beliebiger Zeitreihen nicht von einer zugrundeliegenden Funktion $Y=f(X)$ ausgegangen werden, so daß $h^*(f)$ nicht unbedingt existieren muß. Außerdem ist die getrennte Berechnung beider Entropien sehr zeitaufwendig. Insoweit sind weitere Untersuchungen hierzu notwendig.

Zunächst soll das technische Problem des Rechenaufwands behandelt werden. Für die numerische Berechnung beider Entropien ist eine Nächste-Nachbarn-Suche erforderlich; in einem Fall zur Abschätzung der bedingten Wahrscheinlichkeit $P_\epsilon(y|\mathbf{x})$, im anderen, um eine Funktion lokal anfitten zu können. Hier bietet sich an, beide Berechnungen zu kombinieren.

Für einen beliebigen Referenzzeitpunkt S werden bei vorgegebenem ϵ alle Punktpaare $(\mathbf{x}_{s'}, y_{s'}) \in C_\epsilon^d(\mathbf{x}_S) \times I_\epsilon(y_S)$, sowie alle Punkte $\mathbf{x}_s \in C_\epsilon^d(\mathbf{x}_S)$ gesucht. Wenn diese gefunden sind, ist die Berechnung von $P_\epsilon(y_S|\mathbf{x}_S)$ fast abgeschlossen. Sie ergibt sich durch den Quotienten aus der Anzahl der gefundenen Punktpaare und Punkte. Für eine Berechnung von η_d aus Gleichung (3.6) muß noch eine lineare Gleichung bestimmt werden.

Hierzu kann gemäß den Überlegungen des letzten Kapitels der Punkt (\mathbf{x}_S, y_S) in den Ursprung $(\mathbf{0}, 0)$ verlegt werden, da der konstante Term einer linearisierten Abbildung keinen Einfluß auf den Informationsverlust hat. Dann sind für die Abschätzung einer Ebenengleichung $\tilde{g}(\mathbf{x})$ nur noch die Abstände $\Delta\mathbf{x}_{s'} = \mathbf{x}_{s'} - \mathbf{x}_S$ und $\Delta y_{s'} = y_{s'} - y_S$ von Bedeutung. Sind $m \geq d$ Nachbarpunkte vorhanden, können die Koeffizienten der Ebenengleichung $\tilde{g}(\mathbf{x}) = \sum_{i=1}^d a_i x_i$ durch Lösen eines Gleichungssystems

$$\begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_m \end{pmatrix} = \begin{pmatrix} \Delta x_{1,1} & \Delta x_{1,2} & \cdots & \Delta x_{1,d} \\ \Delta x_{2,1} & \Delta x_{2,2} & \cdots & \Delta x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta x_{m,1} & \Delta x_{m,2} & \cdots & \Delta x_{m,d} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} \quad (4.6)$$

mittels Singulärwertzerlegung bestimmt werden. Die Singulärwertzerlegung liefert im Fall $m > d$ die beste Lösung im Sinne der kleinsten quadratischen Fehler [13].

Durch Einsetzen der Punkte $\mathbf{x}_s \in C^d(\mathbf{x}_S)$ in die Ebenengleichung $\tilde{g}(\mathbf{x})$ kann der Anteil der Punkte $\tilde{\eta}_d$, der durch die gefittete lineare Gleichung auf $I_\epsilon(y_S)$ abgebildet wird, berechnet werden. Da aber nicht von der Existenz einer differenzierbaren Funktion $Y = f(X)$ ausgegangen werden kann, muß $\tilde{\eta}_d$ nicht unbedingt den Wert der bedingten Wahrscheinlichkeit η_d aus (3.6) haben.

Vorerst wird jedoch angenommen, daß durch $\tilde{\eta}_d$ der Anteil der Punkte aus $U_\epsilon(\mathbf{x}_S)$, der durch eine lineare Funktion $g(\mathbf{x}_S)$ auf $I_\epsilon(y_S)$ abgebildet wird, gegeben ist. Dann kann durch

$$\tilde{P}_\epsilon(y_S|\mathbf{x}_S) = \frac{P_\epsilon(y_S|\mathbf{x}_S)}{\tilde{\eta}_d(y_S, \mathbf{x}_S)} \quad (4.7)$$

eine bedingte Wahrscheinlichkeit definiert werden. Die bedingte Wahrscheinlichkeit im Zähler beschreibt den gesamten Informationsverlust, der bei einer Meßgenauigkeit von ϵ auftritt, die bedingte Wahrscheinlichkeit im Nenner jenen Informationsverlust, der durch deterministische Einflüsse hervorgerufen wird. Ist die Ursache des Informationsverlustes rein deterministischer Natur, so nimmt $\tilde{P}_\epsilon(y_S|\mathbf{x}_S)$ den Wert eins an. Mittels dieser bedingten Wahrscheinlichkeit kann eine bedingte Entropie

$$\tilde{H}_\epsilon(Y|X) = -E(\log \tilde{P}_\epsilon(y|\mathbf{x})) \quad (4.8)$$

definiert werden. Die Bildung des Erwartungswertes wird hierbei durch Mittelung von $-\log \tilde{P}_\epsilon(y|\mathbf{x}_S)$ über verschiedene Zeiten S vorgenommen. Der Übergang zu bedingten Dichten bei der Grenzwertbildung für unendlich lange Zeitreihen und verschwindende Umgebungsgrößen liefert

$$\tilde{\rho}(y_S|\mathbf{x}_S) = \frac{\rho(y_S|\mathbf{x}_S)}{\tilde{\eta}_d(y_S, \mathbf{x}_S)} \quad (4.9)$$

und damit die bedingte Entropie

$$\tilde{h}(Y|X) = - \int \log \tilde{\rho}(y|\mathbf{x}) dP(X, Y). \quad (4.10)$$

Ist nun Y durch eine differenzierbare Funktion $f(X)$ bestimmt, so gilt

$$\tilde{h}(Y|X) = h(Y|X) - h^*(f) = 0. \quad (4.11)$$

Damit wäre eine Größe gefunden, die Eigenschaft (2.9) besitzt und somit geeignet ist, Abhängigkeiten in Form von differenzierbaren Funktionen zu quantifizieren.

Wie bereits oben angedeutet, kann in der Regel nicht von einem derartigen Zusammenhang ausgegangen werden. Es können dabei verschiedene Fälle unterschieden werden:

Zunächst soll davon ausgegangen werden, daß Y zwar keine differenzierbare Funktion von X ist, jedoch lokal als eine solche dargestellt werden kann. Dies ist beispielsweise der Fall, wenn y mit gleichen Wahrscheinlichkeiten die Werte $\pm f(x)$ annimmt, das heißt $P(y|x) = 0,5 \quad \forall(x, y)$, und daß lokal sowohl für positive, als auch für negative y eine differenzierbare Funktion f existiert. In diesem Fall sollte die bedingte Entropie den Wert $\log 2$ annehmen. Dies kann durch $\tilde{h}(Y|X)$ nur dann

erreicht werden, wenn die linearen Funktionen $\tilde{g}(\mathbf{x})$ lokal angefitet werden kann, also nur Punktpaare aus $U_\epsilon(\mathbf{x}) \times I_\epsilon(y)$ verwendet werden. Dadurch wird auch gewährleistet, daß im Fall $\tilde{h}(Y|X) = 0$ Unstetigkeitsstellen die Wahrscheinlichkeit null haben.

Eine weitere Möglichkeit ist, daß eine Komponente X_j von X keinen Einfluß auf Y hat. In diesem Fall ist die Wahrscheinlichkeitsverteilung von Y für alle Werte x_j dieser Komponente gleich. Die durch Regression gefundene Abbildung \tilde{g} würde daher für diese Komponente den Koeffizienten $a_j = 0$ haben. Aufgrund der Verwendung der Maximumsnorm haben derartige Komponenten keinen Einfluß auf die bedingte Wahrscheinlichkeit (siehe: Abschnitt 3.3).

Weiterhin kann es vorkommen, daß die in X enthaltene Information nicht ausreicht, um Y vollkommen zu beschreiben; Y kann z.B. von drei Größen abhängen, von denen nur zwei gemessen wurden. In derartigen Fällen kann häufig auch lokal keine lineare Abhängigkeit gefunden werden; für die berechneten Koeffizienten ergibt sich dann $a_i \simeq 0$. Im Gegensatz zum zuvor beschriebenen Fall haben die Komponenten aber einen Einfluß auf $P_\epsilon(y|\mathbf{x})$, so daß die bedingte Entropie kleiner als die Einzelentropie von Y wird. Lediglich der Einfluß der Nichtlinearität kann nicht mehr berechnet werden.

Schließlich muß noch der Fall in Betracht gezogen werden, daß Komponenten von X untereinander abhängig sind, also X_j z.B. eine Funktion von X_i ist. Eine solche Funktion kann lokal wieder linearisiert werden. Damit befinden sich sämtliche Punkte $\mathbf{x}_s \in U_\epsilon(\mathbf{x}_s)$ auf einer $(d-1)$ -dimensionalen Hyperebene; die zu findende Funktion ist überbestimmt. In einer lokalen Hauptachsendarstellung der ϵ -Umgebung besitzt eine der Hauptachsen den Eigenwert null. Wenn der Koeffizient der Ebenengleichung in dieser Richtung verschwindet, ist die Gleichung eindeutig bestimmt. Dies wird durch die Singulärwertzerlegung bewirkt. Die Größe Y hängt nur noch von $(d-1)$ Komponenten ab, die Entropie bleibt bei Vernachlässigung von X_j unverändert. Die in X_j bezüglich Y enthaltene Information ist somit redundant.

4.3 Iteration des Verfahrens

Mit $\tilde{h}(Y|X)$ ist somit eine Größe definiert worden, mit der bestimmt werden kann, ob eine Observable Y vollständig durch andere Größen beschrieben werden kann.

Die modifizierte bedingte Entropie $\tilde{H}_\epsilon(Y|X)$ kann aus Zeitreihen abgeschätzt werden und ist qualitativ – bzgl. der Eigenschaften (2.8) und (2.9) – bei hinreichend langen Zeitreihen und kleinen ϵ eine gute Abschätzung für $\tilde{h}(Y|X)$. Um die am Anfang des Kapitels gestellte Frage, welche Größen am besten für eine Vorhersage geeignet sind, beantworten zu können, wird die vorgestellte Methode zur Berechnung von $\tilde{H}_\epsilon(Y|X_i)$ iterativ angewendet.

Dazu werden zu Beginn nur skalare Zeitreihen $\{y_t\}_{t \leq T}$ und $\{x_{t,i}\}_{t \leq T}$ betrachtet, wobei $i = 1, \dots, N$ die N unterschiedlichen Zeitreihen indiziert. Für jede der N Zeitreihen kann die bedingte Entropie $\tilde{H}_\epsilon(Y|X_i)$ berechnet werden. Je mehr Information in X_i über Y enthalten ist, desto geringer ist die bedingte Entropie. Ist Y unabhängig von X_i , so ist

$$\tilde{H}_\epsilon(Y|X_i) = H_\epsilon(Y) = -E(\log P_\epsilon(y)).$$

Damit ist die oben gestellte Frage, ob in einer Zeitreihe Information über eine andere enthalten ist, bereits beantwortet: Ist

$$\tilde{H}_\epsilon(Y|X_i) < H_\epsilon(Y), \quad (4.12)$$

so ist in X_i Information über Y enthalten. Die meiste Information liefert diejenige Zeitreihe $\{x_{t,I}\}_{t \leq T}$, die den geringsten Wert von $\tilde{H}_\epsilon(Y|X_i)$ ergibt.

Ist $\tilde{H}_\epsilon(Y|X_I) = 0$, so ist die Berechnung abgeschlossen. In diesem Fall kann Y als eine Funktion von X_I dargestellt werden. Andernfalls muß überprüft werden, ob durch Hinzunahme einer weiteren Zeitreihe, die Größe Y besser beschrieben werden kann. Es kann nicht unbedingt davon ausgegangen werden, daß die Zeitreihen, welche im ersten Schritt niedrige bedingte Entropie ergaben, auch die meiste zusätzliche Information über Y bei gegeben X_I beinhalten. Daher werden im zweiten Schritt die Entropien für die durch (X_I, X_i) gegebenen zweidimensionalen Zeitreihen bestimmt. Die Zeitreihen, welche im ersten Schritt keinen Einfluß auf Y aufwiesen, können dabei vernachlässigt werden. Es werden dann diejenigen Meßgrößen gesucht, für die

$$\tilde{H}_\epsilon(Y|(X_I, X_i)) < \tilde{H}_\epsilon(Y|X_I) \quad (4.13)$$

ist. Es gibt, wie im letzten Abschnitt gezeigt, zwei Gründe, weshalb die bedingte Entropie nicht durch Hinzunahme einer weiteren Meßgröße reduziert wird: Entweder liefert diese Meßgröße keine Information, oder aber die Information ist

redundant, also in X_I schon enthalten. Die Zeitreihe $\{x_{t,II}\}_{t \leq T}$, für welche im zweiten Schritt die geringste Entropie berechnet wird, beinhaltet also die meiste zusätzliche, d.h. von X_I unabhängige Information.

Das Verfahren kann iterativ fortgesetzt werden, indem jeweils die Zeitreihe, welche die niedrigste bedingte Entropie liefert, festgehalten wird und die Entropieberechnung in der nächsthöheren Dimension fortgesetzt wird.

Die Berechnung ist dann abgeschlossen, wenn die bedingte Entropie verschwindet oder nicht weiter reduziert wird. Dann sind die Meßgrößen X_I, X_{II}, \dots bestimmt, durch die Y am besten vorhergesagt werden kann.

Ein statisches Beispiel

Die einzelnen Iterationsschritte sollen nun anhand eines Beispiels veranschaulicht werden.

Dazu wird ein statisches Beispiel mit fünf Eingangsgrößen X_1, \dots, X_5 betrachtet, wobei die Zeitreihe $\{x_{t,2}\}$ durch $x_{t,2} = \sin(2\pi\sqrt{x_{t,1}})$ aus der ersten hervorgeht, während die vier anderen Zeitreihen aus unabhängigen, auf $[0, 1]$ gleichverteilten Zufallszahlen bestehen. Die Ausgangsgröße Y ist über

$$y_t = 1,5(x_{t,1} - x_{t,3} + x_{t,3}x_{t,4})$$

aus den Eingangszeitreihen gegeben.

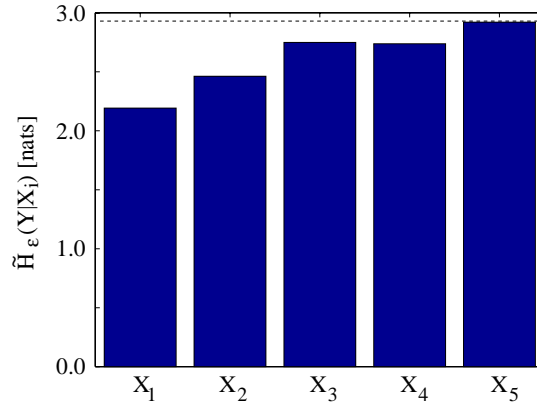


Abbildung 5: Bedingte Entropie $\tilde{H}_\epsilon(Y|X_i)$. Die gestrichelte Linie gibt den Wert der Einzelentropie $\tilde{H}_\epsilon(Y)$ an.

Im ersten Schritt werden nun die bedingten Entropien $\tilde{H}_\epsilon(Y|X_i)$, $i = 1, \dots, 5$ berechnet (Abbildung 5). Dabei wird, wie in der gesamten Arbeit, zur Berechnung

der bedingten Entropien der natürliche Logarithmus verwendet. Die Größe der Entropie wird daher in „nats“ angegeben. In anderen Arbeiten, insbesondere in der Informatik, wird auch der Logarithmus zur Basis 2 verwendet. Dies bedeutet lediglich eine Änderung der Maßeinheiten, welche dann „bits“ sind.

Meßgröße X_5 hat keinen Einfluß auf Y . Die Entropie bzgl. Y wird durch ihre Kenntnis nicht vermindert. Die meiste Information über Y ist in X_1 enthalten. Diese Größe wird daher festgehalten, und im zweiten Schritt werden die bedingten Entropien $\tilde{H}_\epsilon(Y|(X_1, X_i))$ berechnet (Abbildung 6).

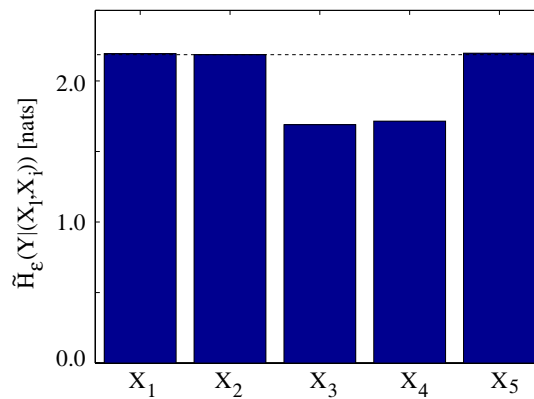


Abbildung 6: Bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, X_i))$. Die gestrichelte Linie gibt den Wert der bedingten Entropie $\tilde{H}_\epsilon(Y|X_1)$ an.

Die Berechnung ergibt, daß durch Hinzunahme der Eingangsgrößen X_1 , X_2 und X_5 keine weitere Reduktion der bedingten Entropie erfolgt. Da die Größe X_5 keinerlei Einfluß auf Y hat, kann sie auch, wie schon im ersten Schritt, keinen zusätzlichen Informationsgewinn liefern. X_1 enthält zwar die meiste Information über Y , jedoch selbstverständlich keinerlei zusätzliche Information, da sie ja von vornherein festgehalten wurde. Ähnlich verhält es sich mit X_2 . Sie lieferte im ersten Schritt die zweitgeringste bedingte Entropie. Allerdings ist X_2 selbst eine Funktion von X_1 . Damit ist die Information, welche X_2 über Y gibt, bereits in X_1 enthalten. Die bedingte Entropie wird also nicht verringert. Tatsächliche neue Information wird nur durch X_3 und X_4 geliefert. Da $\tilde{H}_\epsilon(Y|(X_1, X_3))$ etwas kleiner ist als $\tilde{H}_\epsilon(Y|(X_1, X_4))$, wird im nächsten Schritt zusätzlich zu X_1 noch X_3 festgehalten.

Durch die weitere Hinzunahme von X_4 verschwindet die bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, X_3, X_4))$ nahezu vollständig. Damit ist die Berechnung abgeschlossen.

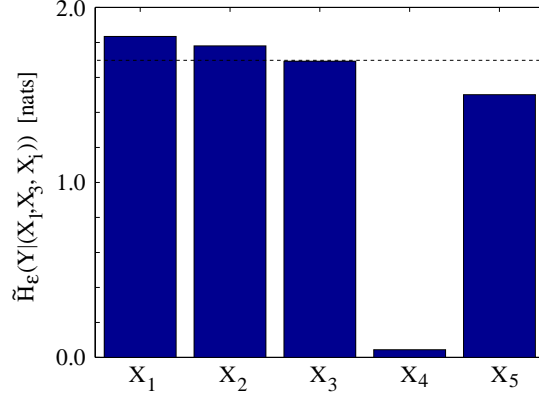


Abbildung 7: Bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, X_3, X_i))$. Die gestrichelte Linie gibt den Wert der bedingten Entropie $\tilde{H}_\epsilon(Y|(X_1, X_3))$ an.

Die Ausgangsgröße Y lässt sich am besten als eine Funktion der Eingangsgrößen X_1 , X_3 und X_4 beschreiben.

Jedoch treten im letzten Schritt bereits deutliche Rechenungenauigkeiten auf: $\tilde{H}_\epsilon(Y|(X_1, X_3, X_4))$ verschwindet nicht vollkommen, und die bedingte Entropie bei den anderen Eingangsgrößen entspricht nicht mehr dem im zweiten Schritt berechneten Wert von $\tilde{H}_\epsilon(Y|(X_1, X_3))$. Diese Fehler lassen sich auf die endliche Anzahl von Zeitreihenpunkten und die nicht verschwindende Umgebungsgröße zurückführen.

4.4 Abhängigkeit der bedingten Entropie $\tilde{H}_\epsilon(Y|X)$ von der Umgebungsgröße und der Länge der Zeitreihe

Im letzten Abschnitt wurde festgestellt, daß $\tilde{H}_\epsilon(Y|X)$ bei hinreichend langen Zeitreihen und genügend kleinen Umgebungsgrößen eine gute Abschätzung für $\tilde{h}(Y|X)$ darstellt. In dem vorgestellten Beispiel ist jedoch zu erkennen, daß die numerisch berechneten Entropien zwar qualitativ die erwarteten Ergebnisse liefern, jedoch quantitativ nicht exakt sind. Daher soll nun der Einfluß, den die Umgebungsgröße ϵ und die Länge der Zeitreihe N_T auf die bedingte Entropie $\tilde{H}_\epsilon(Y|X)$ ausüben, untersucht werden.

Am Beispiel des Henonsystems

$$\begin{aligned} x_{t+1,1} &= 1 - ax_{t,1}^2 + x_{t,2} \\ x_{t+1,2} &= bx_{t,1} \end{aligned} \tag{4.14}$$

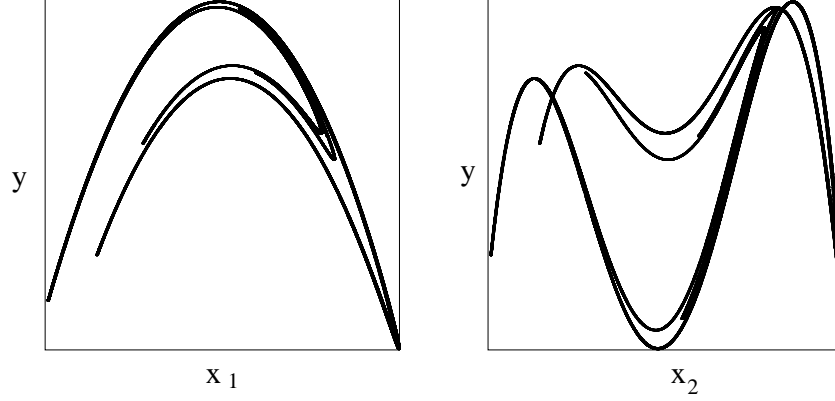


Abbildung 8: Die zeitversetzten Werte y_t bzgl. der x_1 -Komponente über den zeitlich zugehörigen Werten $x_{t,1}$ (links) und $x_{t,2}$ (rechts) des Henonsystems.

mit den üblichen Parametern $a = 1,4$ und $b = 0,3$ wird zunächst der Einfluß von ϵ auf die Berechnung der bedingten Entropie untersucht. Dazu werden als Eingangszeitreihen $\{x_{t,1}\}_{t=1}^T$ und $\{x_{t,2}\}_{t=1}^T$ verwendet, die Ausgangsgröße Y ist durch die zeitversetzten Werte $\{x_{t+1,1}\}_{t=1}^T$ gegeben (Abbildung 8). Die modifizierte bedingte Entropie wird mit der in diesem Kapitel beschriebenen Methode berechnet.

Der Einfluß der Umgebungsgröße

Im ersten Schritt werden wieder die eindimensionalen Zusammenhänge mittels $\tilde{H}_\epsilon(Y|X_i)$ berechnet. Durch den Kehrwert von ϵ ist die Anzahl an Intervallen N festgelegt, welche benötigt wird, um das Intervall $[0, 1]$, auf welches die Zeitreihen normiert sind, abzudecken. In Abbildung 9 sind die bedingten Entropien $\tilde{H}_\epsilon(Y|X_i)$ als Funktion von N dargestellt. Die Kurven wachsen logarithmisch an, $\tilde{H}_\epsilon(Y|X_i)$ kann demnach nicht gegen $\tilde{h}(Y|X_i)$ konvergieren. Dieses Verhalten ist darin begründet, daß Wahrscheinlichkeiten und keine Dichten berechnet werden, welche zur Definition von $\tilde{h}(Y|X_i)$ verwendet wurden. Daher bietet es sich an, $\tilde{H}_\epsilon(Y|X_i)$ im Verhältnis zur Einzelentropie $\tilde{H}_\epsilon(Y)$ (obere Linien in Abbildung 9) zu betrachten.

Wird die bedingte Entropie ohne Berücksichtigung der Korrektur $\tilde{\eta}$, also $H_\epsilon(Y|X) = -E(\log P_\epsilon(y|x))$, berechnet (unterbrochene Linien), so ist bereits eine erhebliche Reduktion der Entropie zu erkennen. Analog zur Definition der

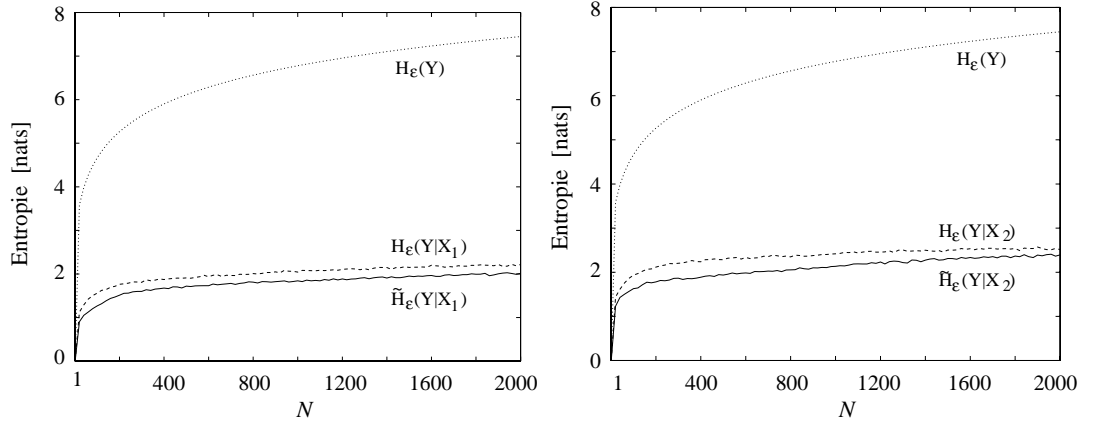


Abbildung 9: Die Entropie $H_\epsilon(Y)$ (gepunktete Linien), sowie die bedingten Entropien $H_\epsilon(Y|X_i)$ (unterbrochene Linien) $\tilde{H}_\epsilon(Y|X_i)$ (durchgezogene Linien) als Funktion von $N = 1/\epsilon$. Links sind die Werte für X_1 und rechts für X_2 zu sehen.

Informationsdimension [14]

$$D_{inf}(Z) = \lim_{\epsilon \rightarrow 0} \frac{H_\epsilon(Z)}{\log(\epsilon)}, \quad (4.15)$$

wobei Z beliebige Systemzustände bezeichnet, kann der Grenzwert

$$D_{cond}(Y|X) = \lim_{\epsilon \rightarrow 0} \frac{H_\epsilon(Y|X)}{\log(\epsilon)} \quad (4.16)$$

als eine „bedingte Informationsdimension“ angesehen werden (Abbildung 10). Sie ist, wenn in X Information über Y enthalten ist, kleiner als die Informationsdimension $D_{inf}(Y)$.

Die durchgezogenen Linien in Abbildung 9 stellen die modifizierte bedingte Entropie dar. Sie ist kleiner als $H_\epsilon(Y|X)$; die Korrektur der nichtlinearen Einflüsse wirkt sich bereits aus. Y ist zwar keine eindeutige Funktion von X_1 oder X_2 , kann aber lokal als eine Funktion von den X_i dargestellt werden (Abbildung 8). Da die Korrekturterme auch nur lokal berechnet werden, führt dies zu einer Verminderung der bedingten Entropie.

Nach dem ersten Schritt wird die Größe X_1 , welche den niedrigeren Wert $\tilde{H}_\epsilon(Y|X)$ ergibt, festgehalten, und die bedingte Entropie wird bei den vollständig gegebenen Systemzuständen (X_1, X_2) berechnet.

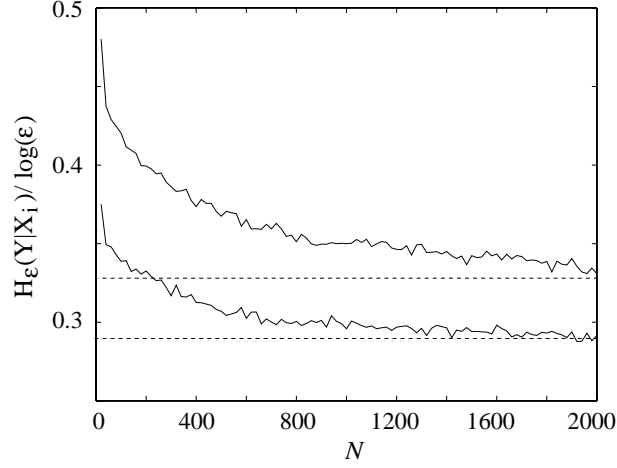


Abbildung 10: Die Quotienten $\frac{H_\epsilon(Y|X_1)}{\log(\epsilon)}$ (unten) und $\frac{H_\epsilon(Y|X_2)}{\log(\epsilon)}$ (oben) konvergieren gegen die jeweilige bedingte Informationsdimension $D_{cond}(Y|X)$ (unterbrochene Linien).

Der Einfluß der Länge der Zeitreihe

Bei feinerer Partitionierung, also größerem N , sinkt zunächst die bedingte Entropie (Abbildung 11). Die geschätzten Werte $\tilde{\eta}(\mathbf{x})$ nähern sich immer mehr den exakten Werten $\eta(\mathbf{x})$ an, so daß die bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, X_2))$, im Rahmen der numerischen Genauigkeit, verschwindet. Wird N allerdings zu groß, so beginnt die bedingte Entropie wieder zu wachsen.

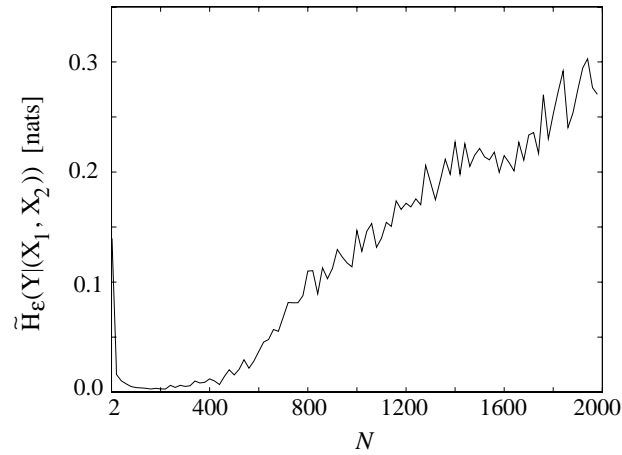


Abbildung 11: Die bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, X_2))$ als Funktion von N .

Die Ursache hierfür liegt in der begrenzten Länge der Zeitreihe, die im Beispiel $N_T = T = 50000$ Datenpunkte beträgt. Die Anzahl der ϵ -Quadrate, die benötigt

wird, um die durch X_1 und X_2 aufgespannte Fläche zu überdecken, ist N^2 . Um eine zweidimensionale Ebene zu bestimmen, werden aber mindestens drei Punkte benötigt. Wird die Anzahl der Punkte in den einzelnen durch X und Y gegebenen ϵ -Würfeln zu klein, kann folglich keine Berechnung des Korrekturterms $\tilde{\eta}$ erfolgen, so daß die berechnete bedingte Entropie wieder anwächst. Bei weiter anwachsendem N ist auch keine sinnvolle Abschätzung von P_ϵ mehr möglich. Allerdings muß in der Regel nicht die gesamte X -Fläche überdeckt werden, sondern nur derjenige Teil, in dem sich auch Datenpunkte befinden. Damit steigt die Anzahl der benötigten ϵ -Würfel nicht mit der euklidischen Dimension, sondern mit der fraktalen Kapazitäts-Dimension D_{Kap} [15]. Daher beginnt das Ansteigen der bedingten Entropie im Beispiel auch erst bei einem Wert von $N \simeq 400$, bei dem die durchschnittliche Anzahl von Punkten in den einzelnen ϵ -Quadraten bereits kleiner als $\frac{1}{3}$ ist.

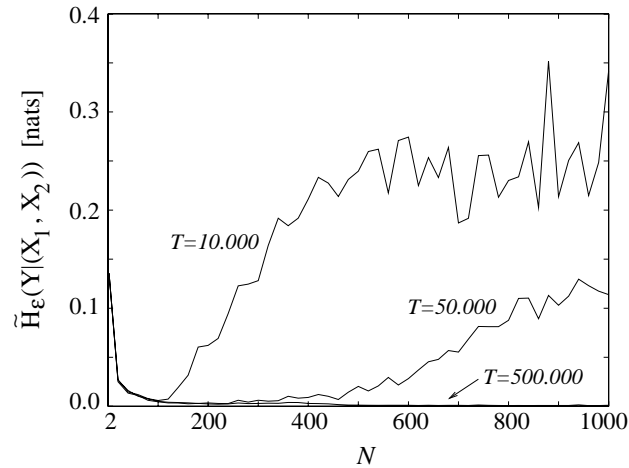


Abbildung 12: Die bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, X_2))$ als Funktion von N bei verschiedenen Zeitreihenlängen. Die Anzahl an Datenpunkten beträgt $T = 500000$, $T = 50000$ und $T = 10000$.

Weitere Berechnungen von $\tilde{H}_\epsilon(Y|(X_1, X_2))$ wurden für Zeitreihen der Länge $T = 500000$, $T = 50000$ und $T = 10000$ durchgeführt (Abbildung 12). Bis zu einem Wert von $N \simeq 100$ verlaufen sämtliche Kurven gleichförmig. Ab $N > 100$ zeigt die Kurve für die kürzeste Zeitreihe deutliche Abweichungen. Die längste Zeitreihe liefert für Werte bis $N = 1000$ bedingte Entropien, die kleiner als 0,002 nats sind, und selbst bei $N = 2000$ ist $\tilde{H}_{0,0005}(Y|(X_1, X_2)) < 0,005$ nats. Für eine Zeitreihe mit 2000 Datenpunkten (nicht dargestellt) wurde bei $N = 60$ dieser

Wert ebenfalls unterschritten. Jedoch traten hier bereits bei $N \simeq 250$ sehr große Fehler in der Berechnung auf.

Die Wahl der richtigen Umgebungsgröße hängt demnach von der Länge der Zeitreihe ab; je kürzer diese ist, desto größer muß ϵ gewählt werden. Andererseits muß ebenfalls gewährleistet sein, daß $\tilde{H}_\epsilon(Y|X)$ noch hinreichend gut abgeschätzt werden kann, also insbesondere die nichtlinearen Einflüsse bestimmt werden können.

Für höherdimensionale Systeme mit einer komplexen Dynamik reichen u.U. eine Million Datenpunkte für eine sinnvolle Abschätzung der bedingten Entropie nicht aus. Zur Identifizierung eindimensionaler Zusammenhänge genügt häufig bereits eine Zeitreihenlänge von $T \simeq 1000$. Da die Wahl von ϵ von der Dynamik des Systems abhängt, kann keine feste Regel dafür angegeben werden; es sollte aber immer so klein gewählt werden, wie es die Länge der Zeitreihe zuläßt.

5 Abhängigkeiten innerhalb einer skalarwertigen Zeitreihe

In diesem Kapitel wird zunächst eine Methode vorgestellt, mit der aus Zeitreihen nichtlinearer dynamischer Systeme Zeitabhängigkeiten ermittelt werden können, die mittels der linearen Standardverfahren nicht identifiziert werden können. Aus diesem Grund werden diese Abhängigkeiten auch häufig als „verborgene Periodizitäten“ bezeichnet. Es wird gezeigt, daß diese zeitlichen Abhängigkeiten auch durch die Berechnung der bedingten Entropie ermittelt werden können. Zusammenhänge zwischen den Verfahren werden untersucht.

5.1 Dichten entlang einer Trajektorie

Ein zeitkontinuierliches, dynamisches System kann gemäß Kapitel 2 durch ein Differentialgleichungssystem $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ und dem zugehörigen Fluß f^t beschrieben werden. Zu einem vorgegebenen Startpunkt \mathbf{x}_0 wird durch $f^t(\mathbf{x}_0)$ eine Trajektorie beschrieben. Zu jedem Punkt dieser Trajektorie kann analog zu Gleichung (4.1) eine Dichte

$$\rho_\epsilon(\mathbf{x}) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{\mathbf{1}_U(f^t(\mathbf{x}_0))}{\mu_L(U_\epsilon(\mathbf{x}))} dt \quad (5.1)$$

abgeschätzt werden. Dabei ist $\mathbf{1}_U$ wieder die Indikatorfunktion zu der Menge $U_\epsilon(\mathbf{x})$. Ortega schlug vor, diese Dichten zur Untersuchung von Zeitreihen nichtlinearer dynamischer Systeme zu verwenden [16, 17].

Folgende Überlegungen waren dafür ausschlaggebend: In einer skalaren Zeitreihe des dynamischen Systems ist Information über die gesamte Dynamik enthalten. Mittels Zeitverzögerungskordinaten [18, 19] kann eine Rekonstruktion der Dynamik vorgenommen werden, d.h. ist eine Größe X des Systems bekannt, so wird ein Zeitverzögerungsvektor

$$\hat{\mathbf{x}}_t = (x_t, x_{t-\Delta\tau}, \dots, x_{t-(d-1)\Delta\tau}), \quad (5.2)$$

d ist dabei die Einbettungsdimension und $\Delta\tau$ die Zeitverzögerung, gebildet, der die Systeminformation enthält³. In dem rekonstruierten System können nun die

³Im folgenden Kapitel wird näher auf die Wahl von d und $\Delta\tau$ eingegangen.

invarianten Dichten $\rho(\hat{\mathbf{x}}) = \rho_\epsilon(\hat{\mathbf{x}})$ bestimmt werden. Damit liegt mit $\rho_t = \rho(\hat{\mathbf{x}}_t)$ wieder eine skalare Zeitreihe vor, welche die Dichten entlang der Trajektorie beschreibt. Die resultierende Zeitreihe enthält Information über die gesamte Dynamik.

Zur Veranschaulichung des Verfahrens soll die x -Komponente des Lorenzsystems

$$\begin{aligned}\dot{x} &= 10(y - x) \\ \dot{y} &= 28x - y - xz \\ \dot{z} &= -\frac{8}{3}z + xy\end{aligned}\tag{5.3}$$

untersucht werden. Dazu wurde die Differentialgleichung numerisch mit einer festen Schrittweite von $t_s = 0,025$ s' integriert⁴. Im ersten Schritt wird dann das Leistungsspektrum bzgl. der x -Komponente berechnet (Abbildung 13).

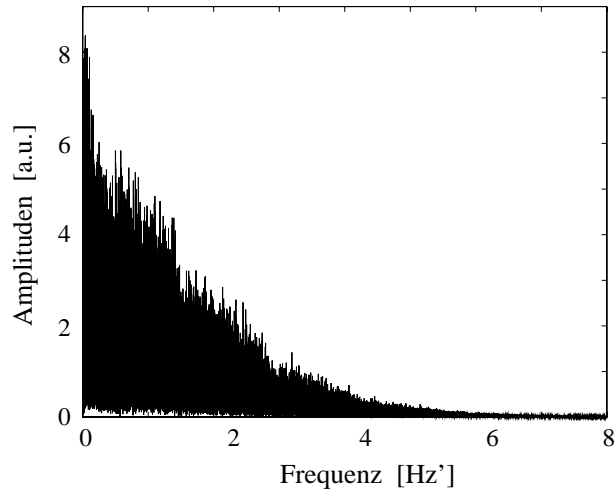


Abbildung 13: Lineares Leistungsspektrum der x -Komponente des Lorenzsystems.

Dieses weist einen deutlichen $(1/f)$ -Abfall auf, und es können ihm somit keine Informationen über dominierende Periodizitäten entnommen werden.

In einem zweiten Schritt wird daher eine Attraktorrekonstruktion gemäß (5.2) durchgeführt. Die verwendeten Parameter sind dabei $d = 3$ und $\Delta\tau = 0,175$ s'. Mittels des in Kapitel 4 beschriebenen Verfahrens werden dann die Dichten ρ_t bestimmt. Die resultierende Zeitreihe der Dichten $\{\rho_t\}_{t=1}^{T'}$ wird daraufhin fouriertransformiert.

⁴Die Einheit s' ist durch einen Integrationsschritt der Länge 1 gegeben; $\text{Hz}' = 1/\text{s}'$.

Deren Leistungsspektrum (Abbildung 14, links) weist ein scharfes lokales Maximum bei $f \simeq 1,2 \text{ Hz}$ auf. Dieses entspricht exakt der dominierenden Frequenz der z -Komponente des Lorenzsystems (Abbildung 14, rechts).

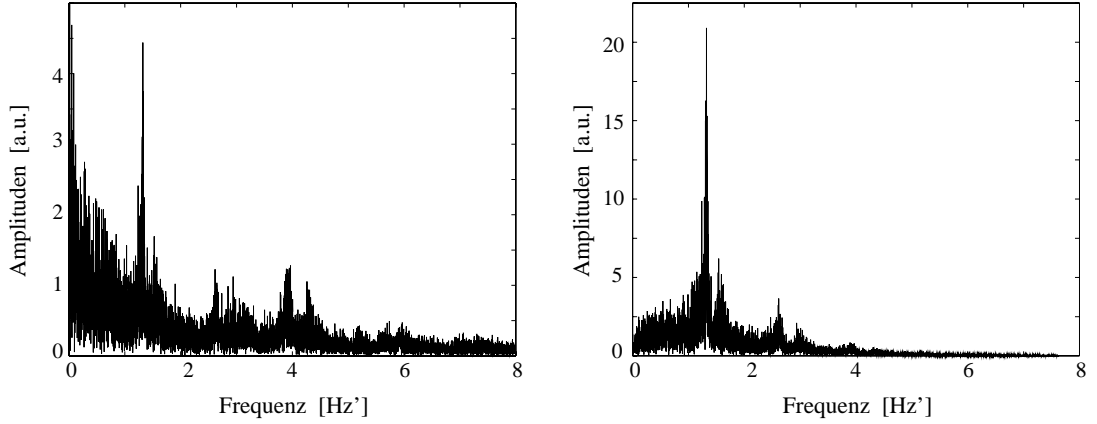


Abbildung 14: Lineare Leistungsspektren der Dichtezeitreihe $\{\rho(\hat{\mathbf{x}}_t)\}$ (links) und der z -Komponente des Lorenzsystems (rechts).

Mit der vorgestellten Methode können also verborgene Periodizitäten, die in der Ausgangszeitreihe mit den linearen Standardmethoden nicht identifizierbar sind, bestimmt werden.

5.2 Zeitentwicklung der Dichten bei iterierten Abbildungen auf einem Intervall

Die Bedeutung der zeitlichen Entwicklung von Dichten entlang einer Trajektorie wird nun anhand eindimensionaler, zeitdiskreter dynamischer Systeme untersucht. Analogien zwischen der oben vorgestellten Methode und dem in dieser Arbeit verwendeten informationstheoretischen Verfahren werden hierbei betrachtet.

Dazu werden iterierte Abbildungen $x_{t+1} = f^1(x_t)$ (2.2) untersucht, wobei $f = f^1$ ein Intervall $I \in \mathbb{R}$ auf sich selbst abbildet. Weiterhin wird vorausgesetzt, daß die invariante Dichte fast überall auf I endlich, also $\rho(x) > 0$ für $x \in I$, ansonsten 0 ist. Dann ist die Verteilungsfunktion [9]

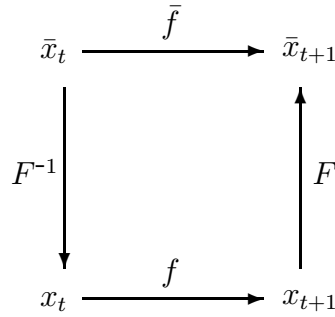
$$F(x) = \int_{-\infty}^x \rho(y) dy \quad (5.4)$$

auf I streng monoton steigend und somit invertierbar.

Im weiteren wird eine Zeitreihe eines solchen Systems untersucht. Diese entsteht durch iterierte Anwendung von f bei einem beliebigen Startpunkt $x_0 \in I$. Werden alle Werte dieser Zeitreihe durch $\bar{x}_t = F(x_t)$ abgebildet, so wird eine neue Zeitreihe $\{\bar{x}_t\}$ gebildet, die eine Gleichverteilung auf dem Intervall $[0, 1]$ erzeugt. Die \bar{x}_t sind Realisierungen eines dynamischen Systems auf dem Einheitsintervall $[0, 1]$, das durch

$$\bar{x}_{t+1} = \bar{f}(\bar{x}_t), \quad \text{mit } \bar{f} = F \circ f \circ F^{-1} \quad (5.5)$$

gegeben ist. Der Zusammenhang der beiden Systeme kann schematisch durch



dargestellt werden. Mittels der Kettenregel kann nun \bar{f} nach \bar{x} abgeleitet werden

$$\frac{d\bar{f}(\bar{x}_t)}{d\bar{x}} = \frac{dF^{-1}(\bar{x}_t)}{d\bar{x}} \cdot \frac{df(x_t)}{dx} \cdot \frac{dF(x_{t+1})}{dx}. \quad (5.6)$$

Da die Dichte gemäß der Definition der Verteilungsfunktion deren Ableitung ist, d.h.

$$\rho(x) = \frac{dF(x)}{dx},$$

kann für die zeitliche Entwicklung der Dichte entlang der Trajektorie der Zusammenhang

$$\rho(x_{t+1}) = \left| \frac{\bar{f}'(\bar{x}_t)}{f'(x_t)} \right| \cdot \rho(x_t) \quad (5.7)$$

abgeleitet werden.

Als Beispiel sei hier die logistische Abbildung

$$f(x) = 1 - 2x^2 \quad (5.8)$$

aufgeführt. Die korrespondierende Abbildung auf dem Einheitsintervall ist die Zeltabbildung

$$\bar{f}(\bar{x}) = \begin{cases} 2x: & 0 \leq x \leq \frac{1}{2} \\ 2 - 2x: & \frac{1}{2} < x \leq 1. \end{cases}$$

Die durch die logistische Abbildung gegebene Dichtefunktion ist [20]

$$\rho(x) = \frac{1}{\pi\sqrt{1-x^2}}.$$

Die Dichte an der Stelle $f(x)$ ist

$$\left| \frac{f'(\bar{x})}{f'(x)} \right| \cdot \rho(x) = \frac{1}{2|x|\pi\sqrt{1-x^2}} = \frac{1}{\pi\sqrt{4x^2-4x^4}} = \frac{1}{\pi\sqrt{1-(1-2x^2)^2}} = \rho(f(x)).$$

Für den eindimensionalen Fall (5.7) wird nun die Autokorrelationsfunktion berechnet. Diese ist als Fouriertransformierte des Leistungsspektrums eine äquivalente Darstellung der zeitlichen Abhängigkeiten.

Mit der Abkürzung $b_{\Delta\tau}(x_t) = \left| \frac{(\bar{f}^{\Delta\tau})'(\bar{x}_t)}{(f^{\Delta\tau})'(x_t)} \right|$ kann die Autokorrelationsfunktion der Dichten als

$$Corr(\Delta\tau) = \frac{E[(\rho_t - E\rho) \cdot (\rho_{t+\Delta\tau} - E\rho)]}{E[(\rho_t - E\rho)^2]} = \frac{E[(\rho_t - E\rho) \cdot (b_{\Delta\tau}(x_t) \cdot \rho_t - E\rho)]}{E[(\rho_t - E\rho)^2]} \quad (5.9)$$

geschrieben werden. Ausmultiplizieren von Zähler und Nenner ergibt

$$Corr(\Delta\tau) = \frac{E[b_{\Delta\tau}(x_t) \cdot \rho_t^2] - (E\rho)^2}{E(\rho_t^2) - (E\rho)^2}. \quad (5.10)$$

Die Korrelationskoeffizienten zu festen Zeitdifferenzen $\Delta\tau$ hängen also weitgehend von den Werten $b_{\Delta\tau}(x_t)$ ab. Der Zähler $|(\bar{f}^{\Delta\tau})'(\bar{x}_t)|$ kann als ein Maß für den Anteil, welchen $\rho(x_t)$ zur Dichte $\rho(x_{t+\Delta\tau})$ beiträgt, angesehen werden. Der Nenner $\left| \frac{1}{(f^{\Delta\tau})'(x_t)} \right|$ entspricht der bedingten Wahrscheinlichkeit η_1 aus (3.6). Er gibt Auskunft über den Informationsverlust, der lokal durch Nichtlinearitäten verursacht wird. Die Autokorrelation der Dichten entlang einer Trajektorie kann damit als Maß für den mittleren Informationsfluß über eine Zeit $\Delta\tau$ betrachtet werden. Ist der Informationsverlust gering, so nimmt die Autokorrelationsfunktion große, ist er jedoch groß, so nimmt sie kleine Werte an.

Aufgrund des Fehlens einer Verteilungsfunktion auf höherdimensionalen Mannigfaltigkeiten kann die obige Herleitung nicht auf allgemeinere Fälle übertragen werden. Dennoch erscheint es nach den bisherigen Überlegungen sinnvoll, den mittleren Informationsverlust über eine Zeit $\Delta\tau$ direkt mittels der bedingten Entropie zu berechnen.

5.3 Bedingte Entropie und Informationsflüsse in skalaren Zeitreihen

Um zeitliche Zusammenhänge zwischen den Zuständen eines Systems bei Messung nur einer skalaren Zeitreihe zu quantifizieren, werden nun die bedingten Entropien $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ berechnet. Es werden also für verschiedene Zeitdifferenzen $\Delta\tau$ die Zeitreihen $\{x_t\}_{t \leq T}$ und $\{x_{t-\Delta\tau}\}_{t \leq T}$ betrachtet. Die eindimensionalen bedingten Entropien ergeben dann eine Funktion der Zeitdifferenzen $\Delta\tau$

$$\tilde{H}_\epsilon(\Delta\tau) = \tilde{H}_\epsilon(X_0|X_{-\Delta\tau}). \quad (5.11)$$

Da bei nichtlinearer Dynamik die Kolmogorov-Sinai-Entropie häufig positiv ist, kann angenommen werden, daß bei größeren Zeitabständen weniger Information über den zukünftigen Zustand vorhanden ist. Damit ist dann ein Ansteigen der Entropiefunktion (5.11) verbunden. Sind jedoch verborgene Periodizitäten in der Zeitreihe vorhanden, so ist bzgl. der entsprechenden Periodenlängen ein größerer Informationsgehalt zu erwarten. Die Kurve der bedingten Entropie sollte dann lokale Minima aufweisen.

Um dies zu überprüfen, wird wieder die x -Koordinate des Lorenzsystems (5.3) untersucht. Dabei werden die in Abschnitt 5.1 eingeführten Zeitreihen verwendet.

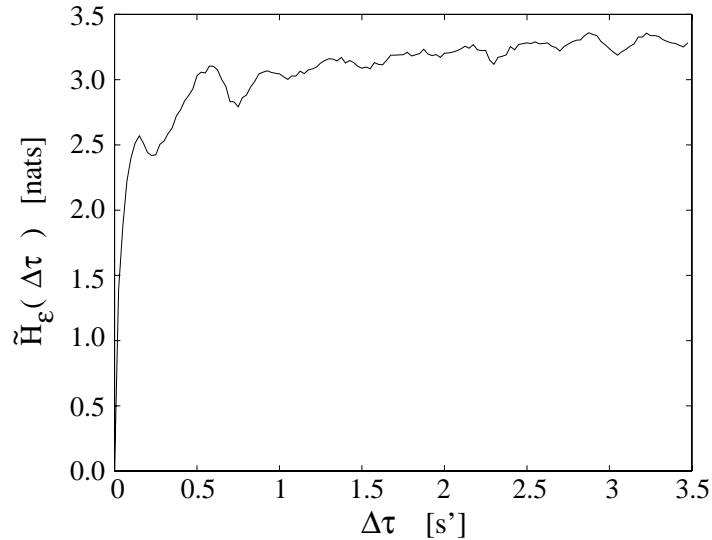


Abbildung 15: Bedingte Entropie $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ als Funktion der Zeitdifferenz $\Delta\tau$ (x -Komponente des Lorenzsystems).

In Abbildung 15 ist die Kurve $\tilde{H}_\epsilon(\Delta\tau)$ der bedingten Entropie dargestellt.

Ohne Zeitverzögerung ($\Delta\tau = 0$) ist die bedingte Entropie 0, da die untersuchten Zeitreihen identisch sind. Mit wachsenden $\Delta\tau$ steigt die bedingte Entropie stark an; während der zeitlichen Entwicklung des Systems geht immer mehr Information verloren. Bei Zeitdifferenzen von $\Delta\tau \simeq 0,25$ s' und Vielfachen von $\Delta\tau \simeq 0,8$ s' besitzt die Funktion lokale Minima.

Zum Vergleich wird nun die Autokorrelation der Dichtezeitreihe aus Abschnitt 5.1 berechnet (Abbildung 16). Entsprechend der im letzten Abschnitt vorgestellten Ergebnisse nimmt sie bei denjenigen Zeitabständen $\Delta\tau$ große Werte an, bei denen im Mittel nach Ablauf dieser Zeit noch viel Information erhalten bleibt.

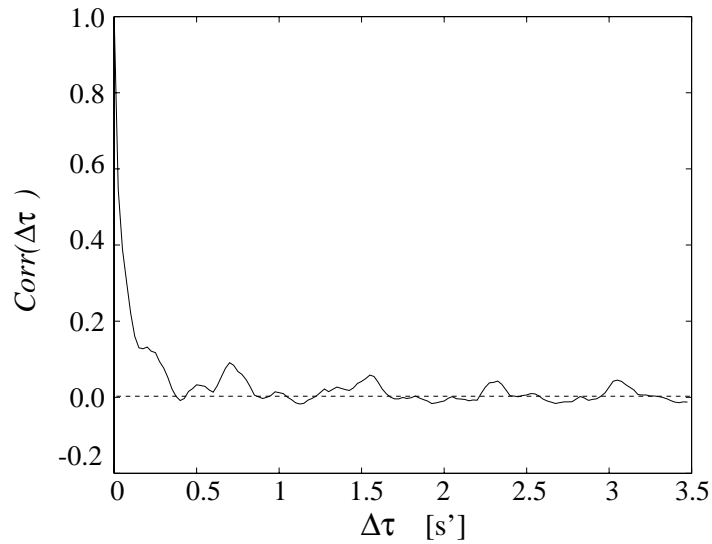


Abbildung 16: Autokorrelationsfunktion der zeitaufgelösten Dichten im aus der x -Komponente rekonstruierten Attraktor des Lorenzsystems.

Aus den Abbildungen 15 und 16 ist deutlich zu erkennen, daß $Corr(\Delta\tau)$ gerade dann hohe Werte annimmt, wenn $\tilde{H}_\epsilon(\Delta\tau)$ lokale Minima aufweist. Das bedeutet, daß mittels der Berechnung der bedingten Entropie aus skalaren Zeitreihen auch verborgene Periodizitäten identifiziert werden können.

Die beiden vorgestellten Methoden dürfen jedoch nicht als äquivalent angesehen werden, da gravierende Unterschiede zwischen beiden bestehen.

Zunächst ist anzumerken, daß die Korrelation ein symmetrisches Maß ist, d.h. $Corr(\Delta\tau) = Corr(-\Delta\tau)$. Damit kann durch sie nur diejenige Menge an Information quantifiziert werden, die sowohl in X_0 als auch in $X_{\Delta\tau}$ enthalten ist. Sie entspricht damit eher der Transinformation (2.10). Die bedingte Entropie gibt

jedoch die in X_0 über spätere Zustände $X_{\Delta\tau}$ enthaltene Information an. Im Falle der logistischen Abbildung aus dem vorangegangenen Abschnitt ist $\tilde{h}(X_1|X_0) = 0$, da X_1 durch die Funktion $f(x) = 1 - x^2$ aus X_0 hervorgeht. Es ist jedoch $\tilde{h}(X_0|X_1) = \log 2$, da jeder Funktionswert genau zwei Urbilder besitzt.

Ein zweiter wesentlicher Unterschied ist, daß durch die Bestimmung der Dichten entlang einer Trajektorie der Informationsfluß im gesamten System untersucht wird. Die bedingte Entropie $\tilde{h}(Y|X)$ gibt jedoch gemäß ihrer Definition an, wieviel Information über eine skalare Größe Y aus der Kenntnis einer anderen Größe X gewonnen werden kann.

6 Rekonstruktion nichtlinearer Dynamik

Im letzten Kapitel wurde gezeigt, daß zeitliche Zusammenhänge zwischen Zuständen aus Zeitreihen dynamischer Systeme durch die Berechnung der bedingten Entropie entdeckt werden können. Dabei wurden nur skalare Zeitreihen des Systems betrachtet. Die Untersuchung des Einflusses von mehrdimensionalen Größen X , die mittels Zeitverzögerungskordinaten gewonnen werden können, führt zur Identifizierung optimaler Parameter für eine Zeitverzögerungseinbettung.

6.1 Bedingte Entropie und die Wahl der Verzögerungszeiten

Die Rekonstruktion chaotischer Attraktoren [18, 19] ist eine der am häufigsten angewandten Methoden bei der Analyse nichtlinearer dynamischer Systeme.

Dazu wird ein Zustand \mathbf{x}_t des dynamischen Systems mit einem aus zeitverzögerten Koordinaten einer einzelnen Systemgröße bestehenden Vektor

$$\hat{\mathbf{x}}_t = (x_t, x_{t-\Delta\tau}, \dots, x_{t-(d-1)\Delta\tau}) \quad (6.1)$$

identifiziert. Gilt für die Einbettungsdimension $d > 2D_{Kap}$, wobei D_{Kap} die Kapazität darstellt, so ist durch die Zeitverzögerungsvektoren (6.1) fast sicher eine Einbettung des Originalattraktors gegeben [19, 21]. Die Wahl der Zeitverzögerung ist bei unendlich langen Zeiten willkürlich, mit der Einschränkung, daß das Originalsystem keine periodischen Bahnen der Länge $\Delta\tau$ und $2\Delta\tau$ enthalten darf. Bei der Analyse endlicher Zeitreihen ist dies nicht mehr der Fall.

Eine Vielzahl von Analysemethoden der nichtlinearen Dynamik basieren auf der Suche nach nächsten Nachbarn. Wie in Kapitel 4 gezeigt, kann bei einer festgehaltenen Anzahl von Datenpunkten die Größe einer ϵ -Umgebung, in der nach Nachbarpunkten gesucht wird, nicht mehr beliebig klein gewählt werden. Die Zeitverzögerung sollte so gewählt werden, daß die Einbettungsdimension möglichst klein gewählt werden kann, da die Anzahl benötigter Datenpunkte bei fester Umgebungsgröße exponentiell mit der Einbettungsdimension ansteigt [22].

Um dies zu erreichen wurden zahlreiche Vorschläge gemacht [23, 24, 25, 26, 27]. Die am häufigsten verwendeten Methoden zielen darauf ab, möglichst unabhängige

ge Koordinaten auszuwählen. Dabei wird davon ausgegangen, daß in ihnen am meisten unabhängige Information enthalten ist und dadurch eine möglichst geringe Einbettungsdimension benötigt wird.

Um diejenigen Zeitverzögerungen herauszufinden, welche unabhängige Koordinaten liefern, werden Autokorrelationsfunktionen berechnet oder die Transinformation (2.10) bestimmt [23, 27]. Letztere hat den Vorteil, daß auch nichtlineare Zusammenhänge berücksichtigt werden.

Zur Berechnung der Transinformation werden zeitversetzte Sequenzen $X = X_0$ und $Y = X_{\Delta\tau}$ einer Zeitreihe betrachtet. Die Elemente der Zeitreihen sind also durch $y_t = x_{t-\Delta\tau}$ miteinander verknüpft. Da die Transinformation eine symmetrische Größe ist, kann das Vorzeichen von $\Delta\tau$ beliebig gewählt werden. Für verschiedene $\Delta\tau$ wird dann die Transinformation

$$I(\Delta\tau) = I(X_{\Delta\tau}, X_0) \quad (6.2)$$

bestimmt, also eine Autotransinformationsfunktion berechnet.

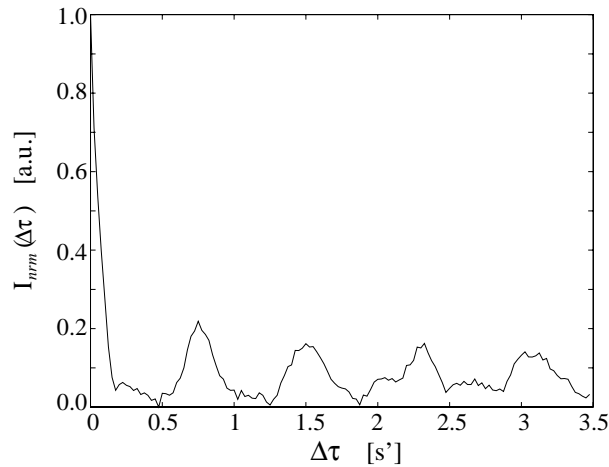


Abbildung 17: Normierte Autotransinformationsfunktion der x -Komponente des Lorenzsystems.

Für die x -Komponente des Lorenzsystems (5.3) wird diese Funktion berechnet; die dazu benötigten Wahrscheinlichkeiten können analog zu der in Kapitel 4 vorgestellten Methode abgeschätzt werden. In Abbildung 17 ist die normierte Funktion zu sehen. Wie schon im letzten Kapitel angesprochen, weist sie deutliche Ähnlichkeiten mit der Autokorrelationsfunktion der Dichtezeitreihen (Abbildung

16) auf. Allerdings sind ihre Maxima deutlicher ausgeprägt, was auf die Verwendung des Logarithmus zurückzuführen ist.

Die Transinformation quantifiziert den gemeinsamen Informationsgehalt zweier Zufallsgrößen, kleine Werte bedeuten also eine große Unabhängigkeit. Damit die $\Delta\tau$ möglichst klein und somit noch viel Information über die Dynamik erhalten bleibt, wird als Zeitverzögerung der Wert gewählt, der das erste Minimum der Autotransinformationsfunktion hervorbringt. Im Beispiel des Lorenzsystems würde also $\Delta\tau \simeq 0,5 \text{ s'}$ gewählt.

Es stellt sich nun die Frage, ob diese Wahl tatsächlich die bestmögliche zur Beschreibung der Dynamik ist. Diese Frage kann durch die Bestimmung der bedingten Entropie beantwortet werden [12]. Wie in Kapitel 4 gezeigt wurde, wird die bedingte Entropie gerade dann klein, wenn möglichst viel unabhängige Information in zwei oder mehr unterschiedlichen Größen über eine andere enthalten ist. Im Fall der Zeitverzögerungseinbettung wird daher der Einfluß von Verzögerungsvektoren $(x_t, x_{t-\Delta\tau})$ auf zukünftige Zustände $x_{t+\Delta t}$ untersucht. Aus diesem Grund wird die bedingte Entropie

$$\tilde{H}_\epsilon^{\Delta t}(\Delta\tau) = \tilde{H}_\epsilon(X_{\Delta t} | (X_0, X_{-\Delta\tau})) \quad (6.3)$$

berechnet. Bei der Zeitverzögerung $\Delta\tau$, die $\tilde{H}_\epsilon^{\Delta t}(\Delta\tau)$ minimiert, kann ein Zustand, der um Δt in die Zukunft versetzt ist, im Mittel am besten durch $(x_t, x_{t-\Delta\tau})$ vorhergesagt werden.

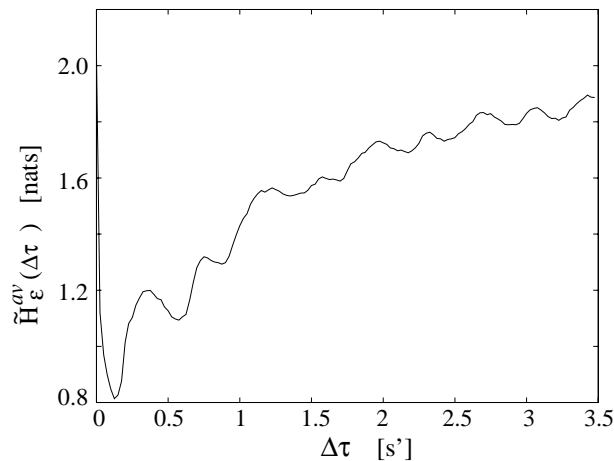


Abbildung 18: Die bedingte Entropie $\tilde{H}_\epsilon^{av}(\Delta\tau)$ der x -Komponente des Lorenzsystems.

Um mögliche Einflüsse von Perioden der Länge $\Delta\tau + \Delta t$ auszuschließen, wird diese bedingte Entropie für verschiedene Vorhersagezeiten Δt berechnet und der Mittelwert

$$\tilde{H}_\epsilon^{av}(\Delta\tau) = \left\langle \tilde{H}_\epsilon^{\Delta t}(\Delta\tau) \right\rangle_{\Delta t} \quad (6.4)$$

gebildet.

Wiederum für die x -Komponente des Lorenzsystems wird $\tilde{H}_\epsilon^{av}(\Delta\tau)$ berechnet (Abbildung 18). Dabei wird die Mittelung (6.4) über Vorhersagezeiten vom ein- bis 25-fachen der Integrationsschrittweite von 0,01 s' durchgeführt. Das Minimum dieser Funktion liegt bei einer Zeitverzögerung von $\Delta\tau \simeq 0,2$ s'. Bei diesem Wert ist einerseits noch viel Information über $X_{\Delta t}$ in $X_{-\Delta\tau}$ enthalten, diese ist aber von der in X_0 enthaltenen weitgehend unabhängig. Mit größer werdenden Zeitverzögerungen steigt auch die bedingte Entropie an, da immer weniger Information über zukünftige Zustände in $X_{-\Delta\tau}$ enthalten ist.

Die Autotransinformationsfunktion hat ihr Minimum bei $\Delta\tau \simeq 0,5$ s', und weist für $\Delta\tau \simeq 0,2$ s' nur ein kleines Nebenminimum auf. Damit kann folgende Schlußfolgerung gezogen werden:

Möglichst große Unabhängigkeit der Verzögerungskordinaten bedeutet nicht zwangsläufig einen großen Informationsgehalt bzgl. der Systemdynamik!

Die Gründe hierfür sollen nun anhand der entsprechenden Attraktorrekonstruktionen erläutert werden (Abbildung 19).

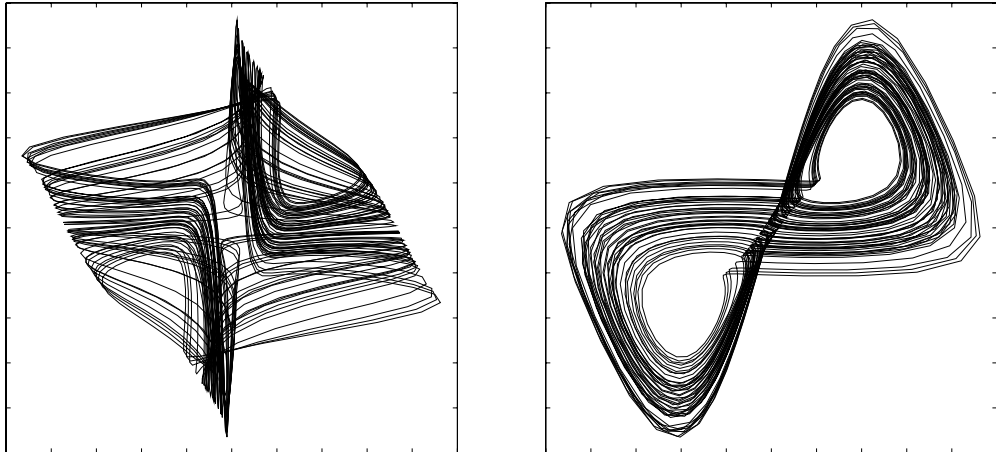


Abbildung 19: Zweidimensionale Rekonstruktionen des Lorenzattraktors mit den durch Transinformation (links) und bedingter Entropie (rechts) ermittelten Verzögerungszeiten.

Wird der Attraktor mittels der durch die Transinformationsfunktion ermittelten Zeitverzögerung rekonstruiert, so sind die einzelnen Punkte möglichst gleichmäßig über einen großen Bereich des Rekonstruktionsraums verteilt. Dies führt jedoch zu einer „Rückfaltung“: Trajektorienabschnitte, die eigentlich zeitlich voneinander getrennt sind, überschneiden sich. Die charakteristische Form des Lorenzattraktors geht verloren (Abbildung 19, links). Die Überschneidung der Trajektorien bedeutet jedoch, daß im rekonstruierten Raum Zustände benachbart sind, die eine zum Teil sehr unterschiedliche weitere zeitliche Entwicklung haben. Damit geht Information über die Dynamik verloren. Genau dieser Punkt ist aber bei der Berechnung der bedingten Entropie $\tilde{H}_\epsilon^{av}(\Delta\tau)$ entscheidend. Da die gemeinsame Information in X_0 und $X_{-\Delta\tau}$ über $X_{\Delta t}$ möglichst gering ist, wird bei der Entfaltung des Attraktors, d.h. bei einer Vergrößerung der Abstände der Teiltrajektorien, die bedingte Entropie zunächst kleiner. Wenn $\tilde{H}_\epsilon^{av}(\Delta\tau)$ ein Minimum erreicht hat, ist der Attraktor optimal entfaltet (Abbildung 19, rechts). Eine Vergrößerung von $\Delta\tau$ führt dazu, daß diese Abstände wieder kleiner werden: Die Information über spätere Zustände wird geringer, $\tilde{H}_\epsilon^{av}(\Delta\tau)$ beginnt wieder zu wachsen. Je größer $\Delta\tau$ wird, desto weniger Information kann den Zeitverzögerungsvektoren entnommen werden. $\tilde{H}_\epsilon^{av}(\Delta\tau)$ konvergiert demzufolge gegen die bedingte Entropie bei ausschließlich gegebenem X_0 .

Die Bestimmung der Verzögerungszeit durch die Berechnung des Minimums von $\tilde{H}_\epsilon^{av}(\Delta\tau)$ hat offenbar deutliche Vorteile gegenüber der Bestimmung mittels Transinformationsfunktionen, da bei der bedingten Entropie die Dynamik des Systems explizit berücksichtigt wird.

6.2 Variable Verzögerungszeiten und die Bestimmung der Einbettungsdimension

Nicht immer zeigt $\tilde{H}_\epsilon^{av}(\Delta\tau)$ ein so deutliches Minimum wie bei dem Lorenzsystem. Dies ist z.B. beim Duffingsystem der Fall. Der Duffing-Oszillator ist ein getriebener Oszillator mit hartem Potential, er wird durch die Differentialgleichung

$$\ddot{x} + D\dot{x} + x + x^3 = c \cdot \cos(\omega t) \quad (6.5)$$

beschrieben. Die hier verwendeten Parameter sind $D = 0,2$, $c = 40$ und $\omega = 1$.

Die Differentialgleichung (6.5) kann als gewöhnliches Differentialgleichungssystem in der Form

$$\begin{aligned}\dot{u} &= v \\ \dot{v} &= c \cdot \cos(\Omega) - Dv - u - u^3 \\ \dot{\Omega} &= \omega\end{aligned}\tag{6.6}$$

geschrieben werden. Die Projektion einer typischen Trajektorie, sowie eine Teilsequenz der Zeitreihe für die u -Komponente sind in Abbildung 20 dargestellt.

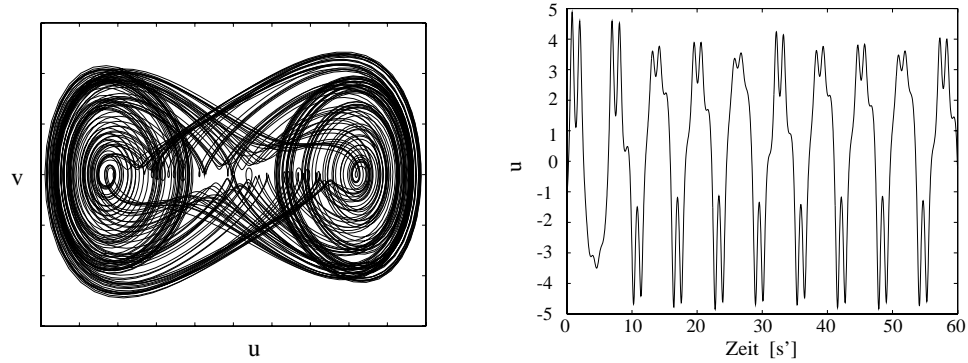


Abbildung 20: Typische Struktur des Duffingattraktors (links). Ausschnitt aus der Zeitreihe der u -Komponente (rechts).

Das System wurde mit einer konstanten Schrittweite von $t_s = \pi/100$ integriert. Die u -Komponente des Systems wird analysiert, d.h. die bedingten Entropien $\tilde{H}_\epsilon^{\Delta t}(\Delta\tau)$ werden berechnet. Die Kurven für die Zeiten $\Delta t = 4t_s, \dots, 12t_s$ sind in Abbildung 21 zu sehen.

Je größer die Vorhersagezeit wird, desto größer wird auch die bedingte Entropie. Dennoch zeigen alle Kurven eine ähnliche Struktur mit mehreren lokalen Minima. Für große Vorhersagezeiten liegt das absolute Minimum bei ca. $\pi/2$ s'. Bei dieser Verzögerungszeit wird die durch die antreibende Sinusschwingung gegebene Dynamik am besten aufgelöst. Für kleine Vorhersagezeiten wird die kleinste Entropie bei $\Delta\tau \simeq 0,3$ s' erreicht. Diese Zeit sorgt für eine bessere Entfaltung der kleineren, höherfrequenten Schwingungsanteile, die in Abbildung 20 (rechts) zu erkennen sind. Dieses Minimum verliert allerdings für größere Δt an Bedeutung, so daß bei Berechnung der gemittelten Entropie $\tilde{H}_\epsilon^{av}(\Delta\tau)$ die Zeitverzögerung $\Delta\tau = \pi/2$ s' als am besten für die Attraktorrekonstruktion geeignet erscheint (Abbildung 22, obere Kurve).

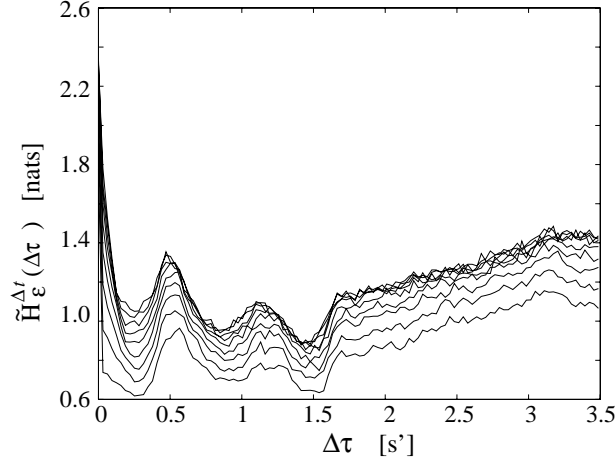


Abbildung 21: Die bedingte Entropie $\tilde{H}_\epsilon^{\Delta t}(\Delta\tau)$ für Vorhersagezeiten $\Delta t = 4t_s$ (untere Kurve) bis $\Delta t = 12t_s$ (oberste Kurve).

Wenn nun aber zwei verschiedene Verzögerungszeiten die Dynamik auf verschiedenen Zeitskalen auflösen, stellt sich die Frage, ob es nicht sinnvoll ist, beide Verzögerungszeiten zur Attraktorrekonstruktion zu verwenden. Die Verwendung verschiedener Verzögerungszeiten widerspricht nicht den Sätzen von Takens und Sauer [18, 19], in denen die Möglichkeit einer Zeitverzögerungseinbettung bewiesen wird.

Um festzustellen, welche dreidimensionale Verzögerungsvektoren die meiste Information über die Systemdynamik enthalten, kann das in Kapitel 4 beschriebene Iterationsverfahren verwendet werden. Dazu wird die im ersten Schritt ermittelte, beste Verzögerungszeit $\Delta\tau_1$ festgehalten und die bedingte Entropie

$$\tilde{H}_\epsilon^{av}(\Delta\tau) = \left\langle \tilde{H}_\epsilon(X_{\Delta t} | (X_0, X_{-\Delta\tau_1}, X_{-\Delta\tau})) \right\rangle \quad (6.7)$$

berechnet (Abbildung 22, mittlere Kurve).

Es ist deutlich zu erkennen, daß die Funktion an der Stelle $\Delta\tau = \Delta\tau_1$ ein Maximum aufweist. Durch die wiederholte Verwendung der ersten Verzögerungszeit kann selbstverständlich keine Information gewonnen werden. An dieser Stelle behält die bedingte Entropie folglich ihren Wert. Im Falle des Duffingsystems ergibt auch die Hinzunahme von $X_{2\Delta\tau}$ nicht den größten Informationsgewinn, sondern die Verwendung einer zweiten Verzögerungszeit $\Delta\tau_2 \simeq 0,3 \text{ s'}$. Dies ist die Zeit, bei der die höherfrequenten Schwingungsanteile am besten aufgelöst werden können. Es erscheint also sinnvoll, verschiedene Verzögerungszeiten zu

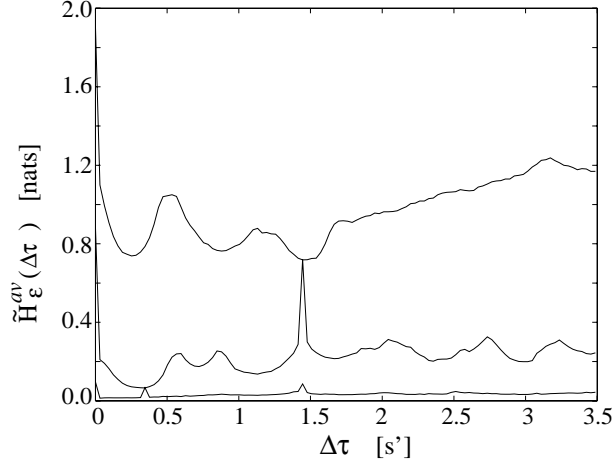


Abbildung 22: Die bedingte Entropie $\tilde{H}_\epsilon^{av}(\Delta\tau)$ für verschiedene Einbettungsdimensionen $d = 2, 3, 4$.

verwenden.

Weiterhin kann durch die Iteration des Verfahrens auch die benötigte minimale Einbettungsdimension abgeschätzt werden.

Eine topologische Einbettung ist eine bijektive, stetige Abbildung, deren Inverse ebenfalls stetig ist [28]. Für Einbettungsdimensionen $d > 2D_{Kap}$ wurde bewiesen, daß durch Bildung von Zeitverzögerungsvektoren mittels (6.1) fast sicher eine Einbettung erhalten wird [19]. Es wird nun gezeigt, daß für eine Rekonstruktion der Dynamik nicht unbedingt äquidistante Verzögerungszeiten notwendig sind. Dazu werden Zeitverzögerungskordinaten

$$X_0, \quad X_{-\Delta\tau_1}, \quad X_{-\Delta\tau_2}, \quad \dots$$

mit beliebigen Verzögerungszeiten $\Delta\tau_i$ vorgegeben. Für derartige Zeitverzögerungsvektoren wird die modifizierte bedingte Entropie $\tilde{H}_\epsilon^{\Delta t}(\Delta\tau)$ analog zu Gleichung (6.3) berechnet. Ist der Grenzwert $\tilde{h}^{\Delta t}(\Delta\tau) = 0$ für mindestens d verschiedene Vorhersagezeiten, welche den Verzögerungszeiten in (6.1) entsprechen, so existiert fast sicher eine stetige Abbildung von dem gegebenen auf einen d -dimensionalen Verzögerungsvektor mit konstanten Verzögerungszeiten, und damit auf den originalen Zustandsvektor. Die vorgegebenen Vektoren können dabei auch eine kleinere Dimension als $d > 2D_{Kap}$ besitzen.

Um nun diese Einbettungsdimension abzuschätzen, wird das oben beschriebene Verfahren solange iteriert, bis der Wert der bedingten Entropie $\tilde{H}_\epsilon^{av}(\Delta\tau)$ im

Rahmen der Rechengenauigkeit, welche von der Länge der Zeitreihe und der Umgebungsgröße ϵ abhängt, verschwindet (Abbildung 22, unterste Kurve). In diesem Fall weist die Kurve der bedingten Entropie keine eindeutigen Minima mehr auf. Im Beispiel des Duffingattraktors ist somit eine dreidimensionale Rekonstruktion mit den berechneten Zeitverschiebungen $\Delta\tau_1 = \pi/2$ s' und $\Delta\tau_2 = 0,3$ s' eine Einbettung.

7 Raum-zeitliche Zusammenhänge

In den vorangegangenen Kapiteln wurden skalare Zeitreihen auf zeitliche Abhängigkeiten untersucht. In diesem Kapitel sollen nun räumliche und raum-zeitliche Zusammenhänge in multivariaten Zeitreihen mittels der Berechnung der bedingten Entropie analysiert werden. Dazu werden zunächst Ketten gekoppelter Abbildungen untersucht. Abschließend wird im zweiten Teil des Kapitels das Verhältnis zwischen treibenden und getriebenen Systemen betrachtet.

7.1 Gekoppelte Abbildungen

Gekoppelte Abbildungen sind ein System von Gleichungen der Form

$$x_{i,t+1} = \sum_{j=1}^N a_{i,j} f_{i,j}(x_{j,t}), \quad i = 1, \dots, N. \quad (7.1)$$

D.h. die Entwicklung einer Größe X_i wird nicht nur von der eigenen Vergangenheit, sondern auch durch „benachbarte“ Größen beeinflusst. Die Koeffizienten $a_{i,j}$ geben die Kopplungsstärken an. Die Größen X_j mit $a_{i,j} \neq 0$ haben über (7.1) einen direkten Einfluß auf die unmittelbare Zukunft von X_i . Die anderen Größen können durch Kopplung mit diesen X_j die weitere Entwicklung von X_i mitbestimmen.

In diesem Abschnitt werden die Zusammenhänge innerhalb einer Kette von gekoppelten Abbildungen anhand des Beispiels

$$\begin{aligned} x_{1,t+1} &= (1-a)\lambda(x_{1,t}^2 - x_1(t)) + a(1-b)\lambda(x_{2,t}^2 - x_{2,t}) \\ x_{i,t+1} &= (1-a)\lambda(x_{i,t}^2 - x_{i,t}) + ab\lambda(x_{i-1,t}^2 - x_{i-1,t}) \\ &\quad + a(1-b)\lambda(x_{i+1,t}^2 - x_{i+1,t}), \quad i = 2, \dots, 7 \\ x_{8,t+1} &= (1-a)\lambda(x_{8,t}^2 - x_{8,t}) + ab\lambda(x_{7,t}^2 - x_{7,t}) \end{aligned} \quad (7.2)$$

untersucht. Die Kopplungsstärke ist dabei durch a und b gegeben, wobei b die Richtung der Kopplung festlegt. Ist $b = 1$, so besteht eine einseitige Kopplung nach links,

$$x_{i,t+1} = (1-a)\lambda(x_{i,t}^2 - x_{i,t}) + a\lambda(x_{i-1,t}^2 - x_{i-1,t}),$$

im Fall $b = 0$ eine einseitige Kopplung nach rechts

$$x_{i,t+1} = (1-a)\lambda(x_{i,t}^2 - x_{i,t}) + a\lambda(x_{i+1,t}^2 - x_{i+1,t}).$$

In den untersuchten Beispielen wird $b = 0, 3$ festgesetzt, so daß das System eine stärkere Kopplung zu höheren Indices aufweist. Der Faktor a wird hingegen von $a = 0$ bis $a = 1$ variiert. Im Fall $a = 0$ liegen acht ungekoppelte Abbildungen

$$x_{i,t+1} = \lambda(x_{i,t}^2 - x_{i,t})$$

vor. Bei der im folgenden festgehaltenen Wahl $\lambda = 2$ haben diese die Struktur der logistischen Abbildung (5.8) auf dem Intervall $[-\frac{1}{2}, \frac{3}{2}]$ und sind somit chaotisch.

7.1.1 Räumliche Zusammenhänge

Zunächst werden strukturelle Zusammenhänge zwischen den einzelnen Meßgrößen untersucht. D.h., es werden nur bedingte Entropien $\tilde{H}_\epsilon(X_j|X_i)$ berechnet; zeitliche Zusammenhänge der Form $\tilde{H}_\epsilon(X_{j,\Delta t}|X_{i,0})$ werden noch nicht analysiert.

Für sämtliche (X_i, X_j) -Kombinationen wird die bedingte Entropie bei verschiedenen Kopplungsstärken a berechnet. Diese Berechnungen ergeben somit für jedes a eine 8×8 -Matrix $\tilde{H}_{i,j}$; in Abbildung 23 sind sie als Oberflächenplot dargestellt.

Die Diagonalelemente nehmen jeweils den Wert $\tilde{H}_{i,i} = 0$ an; eine Größe ist durch sich selbst erwartungsgemäß eindeutig bestimmt. Die Werte der anderen Matrixelemente hängen von der Kopplungsstärke ab.

Bei den nicht gekoppelten Abbildungen ($a = 0$) besteht keinerlei Zusammenhang zwischen den einzelnen Größen. Die bedingten Entropien $\tilde{H}_{i,j}$, mit $i \neq j$, sind also nur durch die Einzelentropien $H_\epsilon(X_i)$ bestimmt (Abbildung 23, oben links).

Mit wachsendem a wird der Einfluß auf benachbarte Kettenglieder größer. Im Oberflächenbild wird die Vertiefung entlang der Diagonalen breiter (Abbildung 23, oben rechts und mitte). Dieser Informationsgewinn erfolgt jedoch nicht symmetrisch. Er wächst mit höheren Indices⁵ i und j . Die Ursache hierfür ist die Wahl von $b = 0, 3$, die eine stärkere Kopplung in Richtung der höheren Indices bewirkt. Ferner ist ansatzweise zu erkennen, daß die Entropiematrix nicht symmetrisch ist, also $\tilde{H}_{i,j} \neq \tilde{H}_{j,i}$ für $i \neq j$. Auch dieser Effekt kann mit der Wahl von b begründet werden.

⁵Hohe Indices sind in Abbildung 23 jeweils vorne links zu sehen.

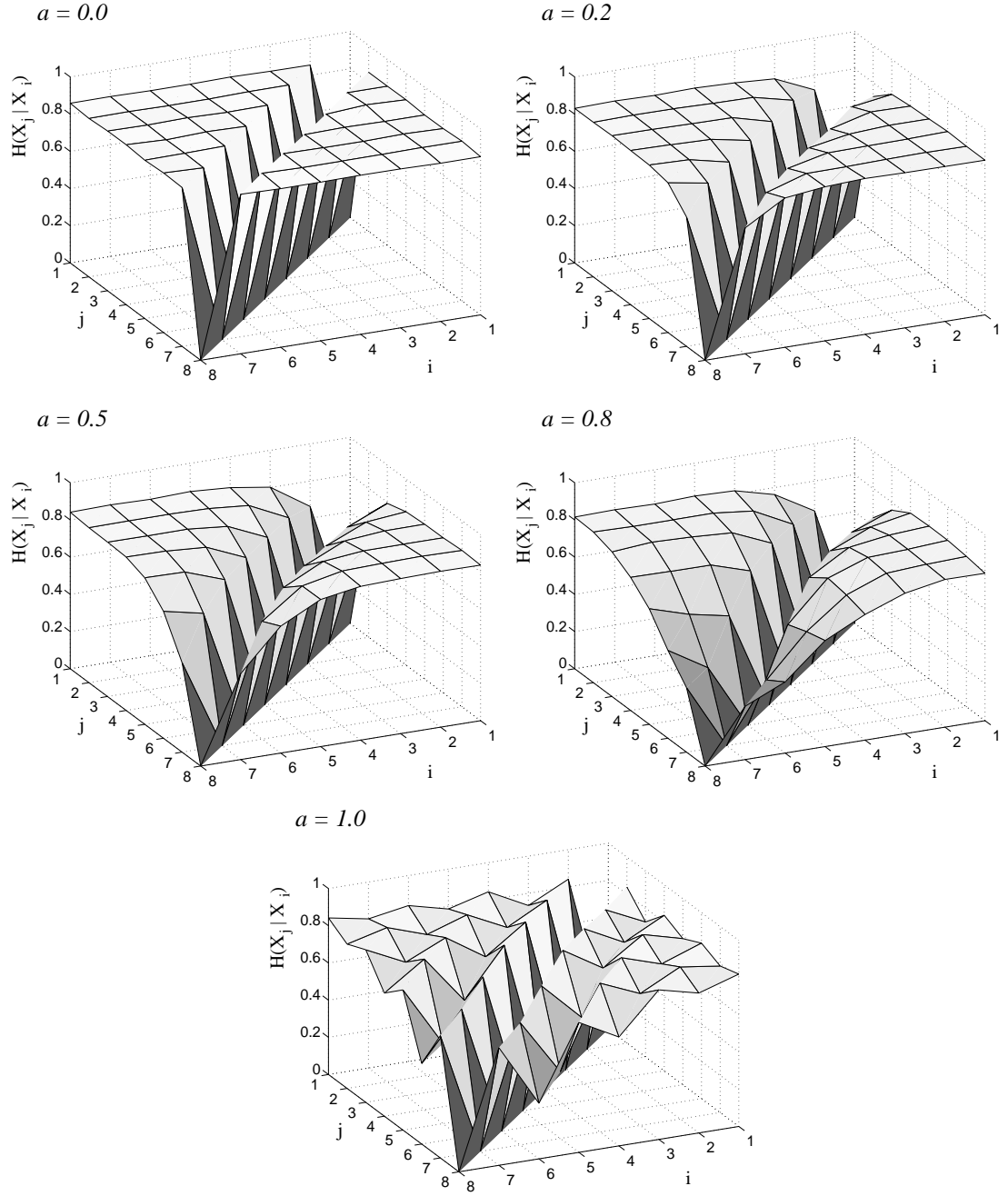


Abbildung 23: Strukturelle Zusammenhänge in der Kette gekoppelter Abbildungen (7.2). Die Kopplungsstärke a wurde von $a = 0$ (oben links) bis $a = 1$ (unten) variiert. Jeder Punkt auf dem 8×8 -Gitter gibt den Wert der bedingten Entropie $\tilde{H}_\epsilon(X_j | X_i)$ an. Dunkle Grauwerte bedeuten niedrige bedingte Entropien, also starke Zusammenhänge.

Hervorzuheben ist, daß die beschriebenen, strukturellen Zusammenhänge nicht direkt aus dem Gleichungssystem (7.2) abzuleiten sind. Die einzelnen Zeitreihenelemente $x_{i,t}$ und $x_{j,t}$ zu gleichen Zeiten t unterliegen nicht dieser Abbildungsvorschrift. Dennoch wird durch die Dynamik die Struktur des Systems derart beeinflußt, daß Zusammenhänge identifiziert werden können. Ein interessanter Fall ergibt sich für $a = 1$ (Abbildung 23, unten). In diesem Fall hängt die direkte Zukunft einer Größe X_i zur Zeit $t + 1$ nur noch von den Zuständen $x_{i-1,t}$ und $x_{i+1,t}$ der benachbarten Kettenglieder ab, jedoch nicht mehr vom eigenen Zustand $x_{i,t}$. Der Einfluß auf die eigenen Zustände erfolgt zeitversetzt nach $\Delta t = 2$. Bei dieser Zeitdifferenz sind die Zustände der nächsten Nachbarn ohne Bedeutung. Die übernächsten Kettenglieder haben eine gemeinsame Vergangenheit mit X_i , während die direkten Nachbarn aufgrund der zeitlich alternierenden Abhängigkeiten diese nicht besitzen. Dies führt zu der im Oberflächenbild erkennbaren, wellenartigen Struktur. In den Größen X_i ist viel Information über die jeweils übernächsten Kettenglieder enthalten. Werden die Zustände der übernächsten Nachbarn gegeneinander aufgetragen, ist eine klare Struktur in der von ihnen aufgespannten Ebene zu erkennen (Abbildung 24, links). Dies ist bei direkt benachbarten Größen nicht der Fall (Abbildung 24, rechts).

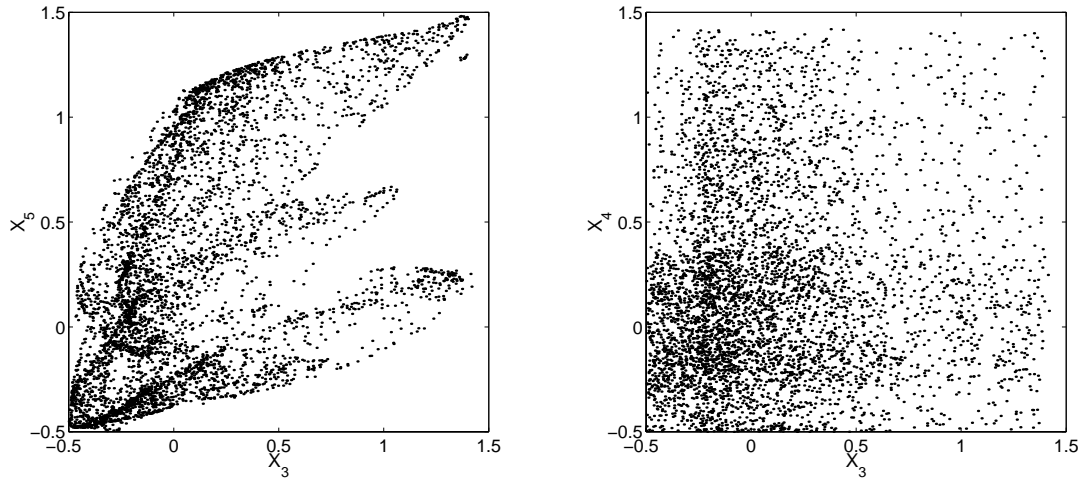


Abbildung 24: Zustandsraumprojektionen auf die X_3 - X_5 -Ebene (links) und die X_3 - X_4 -Ebene (rechts).

Derartige strukturelle Zusammenhänge können bis zur Synchronisation [29, 30] führen, d.h. es existiert dann eine Funktion $X_i = g(X_j)$. In diesem Fall ver-

schwindet die bedingte Entropie. Es ist somit insbesondere möglich, Synchronisation zwischen Zeitreihen zu erkennen, die nicht direkt aus der formalen Beschreibung der Dynamik hervorgeht.

7.1.2 Raum-zeitliche Zusammenhänge

Bislang wurden lediglich räumliche Zusammenhänge analysiert. Im folgenden soll die Methode zur Identifizierung von raum-zeitlichen Abhängigkeiten verwendet werden. Dazu werden zeitverschobene Zeitreihen untersucht und deren bedingte Entropien $\tilde{H}_\epsilon(X_{j,\Delta t}|X_{i,0})$ berechnet.

Abbildung 25 zeigt wiederum Oberflächendarstellungen der Matrizen $\tilde{H}_{i,j}(\Delta t)$. Diese sind für eine Kopplungstärke von $a = 0,5$ berechnet; wie im letzten Abschnitt ist $b = 0,3$. Im Fall $\Delta t = 0$ (oben links) entspricht die Abbildung der aus 23 (Mitte links).

Mit größer werdender Zeitdifferenz Δt wird die Information, die in einzelnen Größen über die Zukunft enthalten ist, geringer. Die Iteration von (7.2) zeigt, dass pro Iterationsschritt ein oder zwei Größen zusätzlich benötigt werden, um einen zukünftigen Zustand vorherzusagen. Für $\Delta t > 7$ werden alle acht Kettenglieder benötigt, um eine Größe $X_{i,\Delta t}$ exakt zu bestimmen.

Die durch b hervorgerufene Asymmetrie tritt bei der Berechnung von $\tilde{H}_{i,j}(\Delta t)$ deutlich hervor. Die Linie minimaler bedingter Entropie verschiebt sich in Richtung höherer Indices, also bei der gewählten Perspektive nach links.

An den Kettenenden X_1 und X_8 ist die bedingte Entropie am geringsten. Da diese nur einseitig gekoppelt sind, werden weniger Größen benötigt, um kurzzeitige Vorhersagen zu treffen, als bei den mittleren Kettengliedern. Daher ist der Informationsverlust in einzelnen Größen während der zeitlichen Entwicklung geringer. Eine weitere Asymmetrie tritt an den Kettenenden auf: So ist bei einer Zeitdifferenz von $\Delta t = 1$ (Abb. 25, oben rechts) die bedingte Entropie $\tilde{H}_\epsilon(X_{8,1}|X_{7,0})$ fast so gering wie die bedingte Entropie bzgl. der eigenen Vergangenheit $\tilde{H}_\epsilon(X_{8,1}|X_{8,0})$, während $\tilde{H}_\epsilon(X_{7,1}|X_{8,0})$ erheblich größer ist. Bei größeren Zeitabständen verstärkt sich dieser Effekt noch, dann ist $\tilde{H}_\epsilon(X_{8,\Delta t}|X_{7,0}) < \tilde{H}_\epsilon(X_{8,\Delta t}|X_{8,0})$ (Mitte links bis unten rechts). Dies ist wiederum darin begründet, daß die Kette nicht zyklisch fortgesetzt ist, die Kettenglieder X_1 und X_8 also nur mit einer weiteren Größe gekoppelt sind. Es sind daher nur zwei Größen notwen-

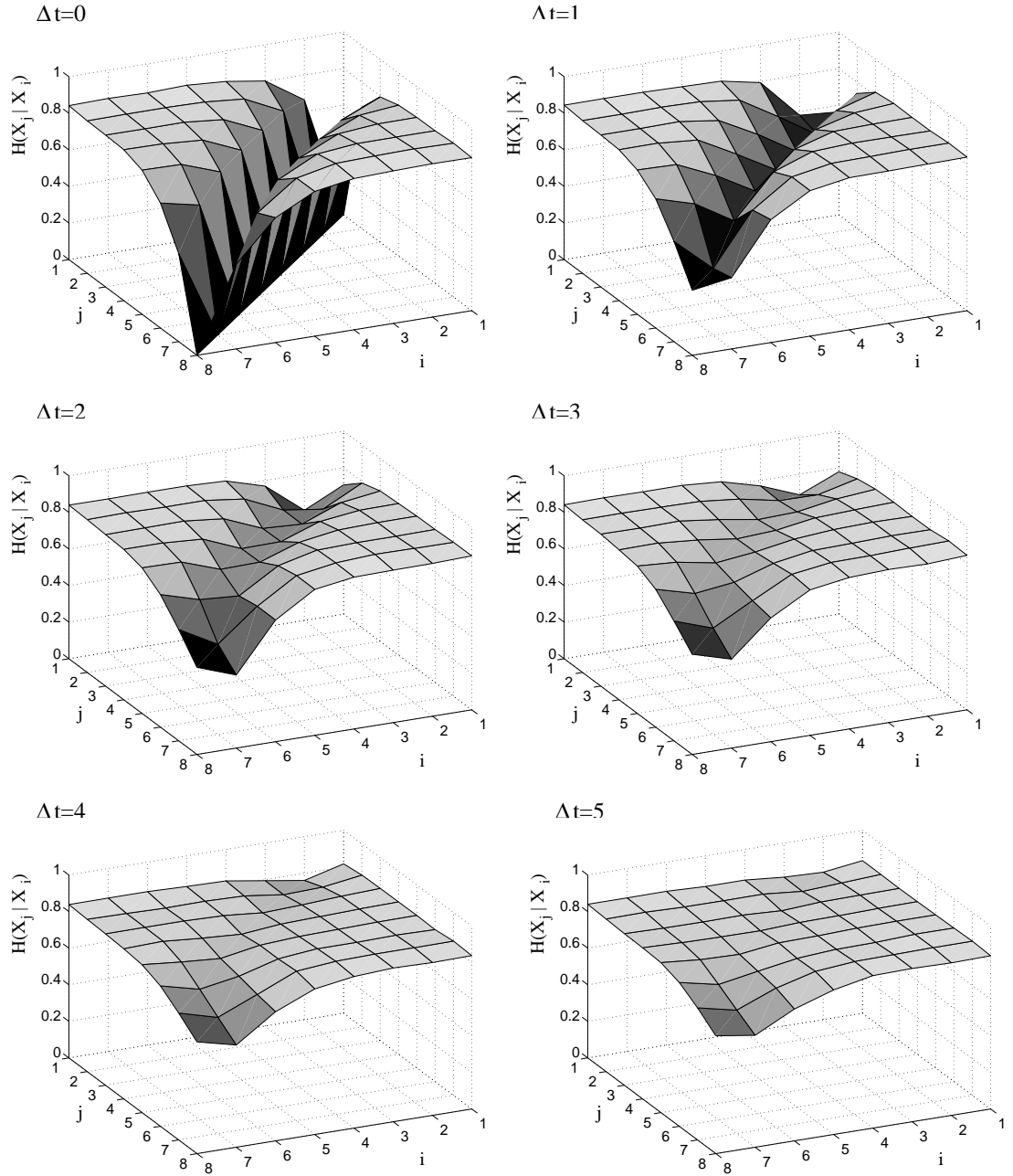


Abbildung 25: Raum-zeitliche Zusammenhänge in der Kette gekoppelter Abbildungen (7.2). Die Kopplungsstärke beträgt $a = 0,5$. Der zeitliche Abstand Δt wird von $\Delta t = 0$ (oben links) bis $\Delta t = 5$ (unten rechts) variiert. Jeder Gitterpunkt gibt den Wert der bedingten Entropie $\tilde{H}_\epsilon(X_{j,\Delta t}|X_{i,0})$ an. Dunkle Grauwerte bedeuten niedrige bedingte Entropien, also starke Zusammenhänge.

dig, um den zukünftigen Zustand von X_1 oder X_8 zu beschreiben, während für die anderen X_i drei Größen notwendig sind. Mit fortschreitender Zeit gewinnen auch entferntere Nachbarn an Bedeutung. Deren Information ist aber bereits teilweise in X_2 bzw. X_7 eingegangen. Damit kann der Informationsverlust über die Kettenenden in diesen Größen geringer sein als in den Kettenenden selbst.

Der Einfluß der benachbarten Kettenglieder wächst auch mit zunehmender Kopplungsstärke a . In Abbildung 26 ist der Einfluß der Nachbarn auf X_1 durch die bedingten Entropien $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{i,0})$ bei gegebenen $X_{i,0}$ für $a = 0,0$ bis $a = 1,0$ graphisch dargestellt.

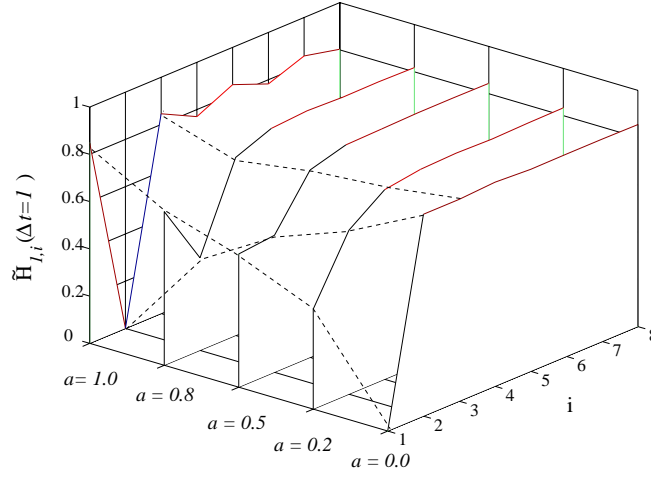


Abbildung 26: Bedingte Entropien $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{i,0})$ bei verschiedenen Kopplungsstärken a .

Im ungekoppelten Fall ($a = 0$) sind zukünftige Zustände von X_1 eindeutig durch den gegenwärtigen Zustand $x_{1,0}$ festgelegt. Die modifizierte bedingte Entropie⁶ beträgt für alle Zeitabstände Δt im Rahmen der Rechengenauigkeit $\tilde{H}_\epsilon(X_{1,\Delta t}|X_{1,0}) = 0$. Im dargestellten Fall $\Delta t = 1$ wächst sie mit steigender Kopplungsstärke, bis bei $a = 1$ kein Einfluß mehr festzustellen ist. Bei $a = 1$ ist $X_{1,1}$, wie durch Iteration von (7.2) erkenntlich wird, eindeutig durch $X_{2,0}$ bestimmt, d.h. $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{2,0}) = 0$. Im Fall $\Delta t = 1$ sind die bedingten Entropien $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{1,0})$ und $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{2,0})$ bzgl. der Kopplungsstärke weitgehend zueinander symmetrisch. Diese Symmetrie bleibt bei größeren Δt jedoch nicht mehr

⁶Dies gilt nicht für die „gewöhnliche“ Entropie $h(X_{1,\Delta t=1}|X_{i,0})$ (4.4), welche nichtlineare Einflüsse nicht berücksichtigt.

erhalten. Während für $a = 0$ die Entropie $\tilde{H}_\epsilon(X_{1,\Delta t}|X_{1,0}) = 0$ bleibt, wird die Zukunft von X_1 bei $a = 1$ von immer mehr Größen beeinflusst; die bedingte Entropie wird größer. Dies geschieht jedoch nicht zeitlich monoton, sondern alternierend in Zweisritten.

Diese alternierenden Einflüsse wurden bereits bei der Berechnung der rein räumlichen Zusammenhänge sichtbar (Abbildung 23, unten). Auch bei den hier berechneten raum-zeitlichen Zusammenhängen treten sie wieder auf. Auch andere räumliche Strukturen, wie die Erweiterung des Einflußbereichs, sind wiederzuerkennen. Sowohl strukturelle als auch formale funktionale Zusammenhänge haben demnach Auswirkungen auf die modifizierte bedingte Entropie.

Es stellt sich nun die Frage, ob die durch (7.2) gegebenen Abhängigkeiten anhand der Zeitreihe identifiziert werden können. Dazu wird bei einer Zeitdifferenz vom $\Delta t = 1$ die Größe X_1 bei einer Kopplungsstärke von $a = 0,5$ untersucht. Die Analyse erfolgt mittels der in Abschnitt 4.3 vorgestellten Methode.

Zunächst werden die Entropien $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{i,0})$ für alle i berechnet. Das Ergebnis ist in Abbildung 27 als Säulendiagramm dargestellt.

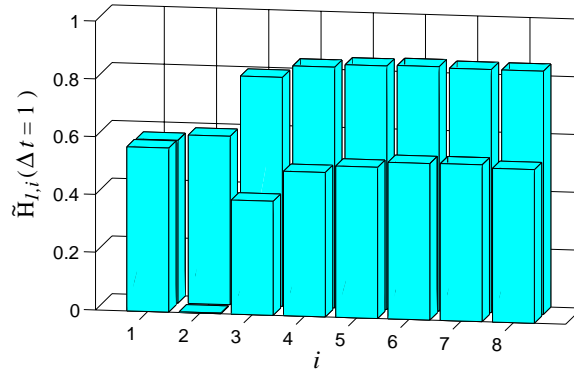


Abbildung 27: Bedingte Entropien $\tilde{H}_\epsilon(X_{1,\Delta t=1}|X_{i,0})$ (hintere Reihe) und $\tilde{H}_\epsilon(X_{1,\Delta t=1}|(X_{1,0}, X_{i,0}))$ (vordere Reihe) bei einer Kopplungsstärke von $a = 0,5$.

Die Größe $X_{I,0}$, welche die geringste bedingte Entropie ergibt, wird festgehalten und anschließend die bedingte Entropie $\tilde{H}_\epsilon(X_{1,\Delta t=1}|(X_{I,0}, X_{i,0}))$ berechnet. Das Ergebnis dieser Berechnung ist durch die vordere Säulenreihe in Abbildung 27 dargestellt: Nachdem im ersten Schritt $X_{1,0}$ den größten Informationsgewinn erbrachte, verschwindet im zweiten Schritt bei Hinzunahme von $X_{2,0}$ die bedingte Entropie $\tilde{H}_\epsilon(X_{1,\Delta t=1}|(X_{1,0}, X_{2,0}))$ fast völlig. Dies bedeutet, daß diese beiden

Größen alleine für den Zustand von X_1 nach einem Iterationsschritt verantwortlich sind, was auch tatsächlich Gleichung (7.2) entspricht.

Es ist also prinzipiell möglich, Informationen über die Dynamik eines Systems durch die Berechnung modifizierter bedingter Entropien zu erhalten. Für die Identifizierung komplexer, höherdimensionaler Abhängigkeiten, also insbesondere solchen bei größeren Zeitdifferenzen werden nach den Ergebnissen aus Abschnitt 4.4 jedoch sehr lange Zeitreihen benötigt.

Dies ist ein numerisches Problem, das durch den Algorithmus bedingt ist. Es besteht jedoch auch ein prinzipieller Unterschied zu der Analyse eines statischen Systems wie dem aus Abschnitt 4.3. Die Abhängigkeiten in statischen Systemen sind eindeutig durch die gegebenen Abbildungen festgelegt. In den vorangegangenen Unterabschnitten wurde jedoch gezeigt, daß in dynamischen Systemen auch strukturelle Zusammenhänge auftreten; auch die relativ kleinen Entropiewerte bei gegebenen $X_{3,0}$ in Abbildung 27 sind kein numerisches Artefakt. Sind die strukturellen Zusammenhänge größer als die durch die eigentliche Dynamik bestimmten Abhängigkeiten, wie es z.B. bei Synchronisation der Fall ist, so liefert die verwendete Methode nicht mehr zwangsläufig die richtigen Ergebnisse. Zwar verschwindet die modifizierte bedingte Entropie $\tilde{h}(X_{i,\Delta t}|X_{*,0})$, wenn in X_* alle für die Dynamik wesentlichen Größen enthalten sind; umgekehrt kann jedoch nicht aus $\tilde{h}(X_{i,\Delta t}|X_{*,0}) = 0$, für alle $X_i \in X_*$, geschlossen werden, daß die Dynamik durch X_* vollständig bestimmt wird. Ein Ursache-Wirkung-Verhältnis ist nur eine hinreichende Bedingung für das Verschwinden der modifizierten bedingten Entropie; deren Verschwinden alleine impliziert jedoch noch keinen derartigen Determinismus.

7.2 Getriebene Systeme

In diesem Abschnitt werden nun zeitkontinuierliche Systeme betrachtet. Die zeitliche Entwicklung einer Größe in einem derartigen System ist immer von ihrem gegenwärtigen Zustand abhängig. Besonders bei kurzen Zeitabständen ist der Einfluß anderer Systemgrößen vergleichsweise gering. Es lassen sich zwar weiterhin raum-zeitliche Zusammenhänge identifizieren, die formalen funktionalen Abhängigkeiten sind aber kaum noch zu rekonstruieren. Um dennoch Aussagen über funktionale Zusammenhänge machen zu können, werden zeitliche Struktura-

ren wie in den Kapiteln 5 und 6 analysiert. Diese Methode wird am Beispiel des Lorenz-getriebenen Duffingsystems demonstriert.

Dieses entspricht dem Duffingsystem (6.5), mit der Ausnahme, daß es anstelle einer Kosinusschwingung von der x -Komponente des Lorenzsystems (5.3) getrieben wird. Es läßt sich folglich als fünfdimensionales System gewöhnlicher Differentialgleichungen beschreiben. Es ist

$$\begin{aligned}\dot{x} &= k(10(y-x)) \\ \dot{y} &= k(28x - y - xz) \\ \dot{z} &= k\left(-\frac{8}{3}z + xy\right) \\ \dot{u} &= v \\ \dot{v} &= c'x - \frac{1}{5}v - u - u^3,\end{aligned}\tag{7.3}$$

wobei das hier gewählte $k = 0,2$ die Dynamik des Lorenzsystems (x, y, z) verlangsamt; die Kopplungsamplitude zum Duffingsystem (u, v) ist durch $c' = 1,4$ gegeben.

Zunächst werden die zeitlichen Strukturen innerhalb der x - und u -Komponenten des Systems untersucht. Das bedeutet, die bedingten Entropien $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ und $\tilde{H}_\epsilon(U_0|U_{-\Delta\tau})$ werden wie in Kapitel 5 berechnet (Abbildung 28).

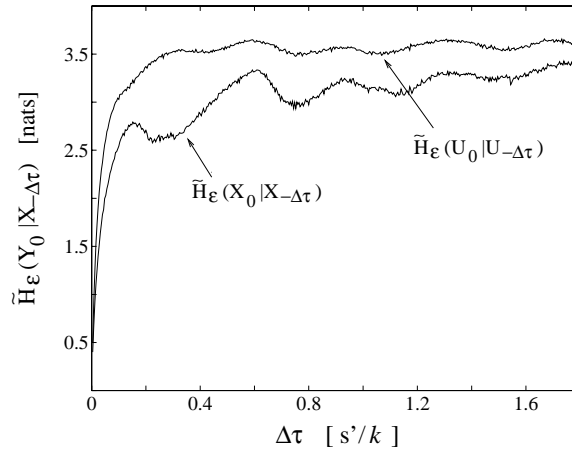


Abbildung 28: Bedingte Entropien $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ (untere Kurve) und $\tilde{H}_\epsilon(U_0|U_{-\Delta\tau})$ (obere Kurve).

Das Lorenzsystem ist von der Dynamik des Duffing-Oszillators vollkommen unabhängig, folglich hat die Entropiekurve $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ exakt die bereits vorgestellte Struktur. Die u -Komponente ist andererseits über \dot{v} in Gleichung (7.3)

vom Lorenzsystem beeinflusst. Abgesehen von Unterschieden für kleine Zeiten $\Delta\tau$ besitzt die Funktion $\tilde{H}_\epsilon(U_0|U_{-\Delta\tau})$ die gleichen lokalen Extrema wie $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$. Der Verlauf der Entropiekurven liefert folglich einen ersten Hinweis auf einen Zusammenhang zwischen beiden Systemen.

Allerdings kann auf diese Weise kein direkter Bezug zwischen einzelnen Größen hergestellt werden, da diese unabhängig voneinander analysiert werden. Um direkte Zusammenhänge zu detektieren, werden daher bedingte Entropien zwischen verschiedenen Größen bei unterschiedlichen Zeitabständen berechnet, also in diesem Fall $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$. Es werden somit raum-zeitliche Strukturen analysiert.

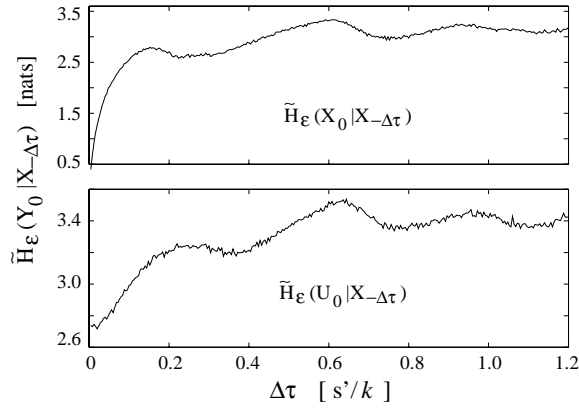


Abbildung 29: Bedingte Entropien $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ (oben) und $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$ (unten).

Hier zeigt sich, daß die in X über zukünftige Zustände von U enthaltene Information fast die gleichen zeitlichen Abhängigkeiten aufweist wie die Information über die eigenen zukünftigen Zustände (Abbildung 29). Das bedeutet die dynamische Struktur des Lorenzsystems wird auf den Duffing-Oszillator übertragen. Dabei geht allerdings Information verloren und die modifizierte bedingte Entropie wird größer. Außerdem ist eine Zeitverzögerung in der Informationsübertragung auszumachen. Die Extrema der Entropiefunktion sind in Richtung längerer Zeiten verschoben.

Weiterhin kann der Informationsgehalt der y -Komponente über die zeitliche Entwicklung der u -Komponente untersucht werden (Abbildung 30, unten). Die bedingte Entropie wird zunächst mit wachsenden Zeitabständen geringer. Nachdem sie ihr absolutes Minimum erreicht hat, beginnt sie wieder zu wachsen. Der anschließende Kurvenverlauf entspricht wieder dem von $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$. Dies ist

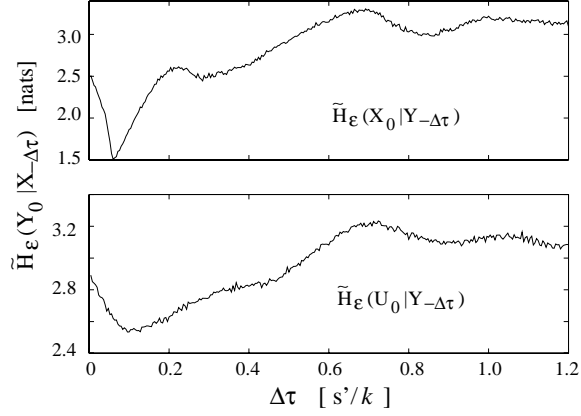


Abbildung 30: Bedingte Entropien $\tilde{H}_\epsilon(X_0|Y_{-\Delta\tau})$ (oben) und $\tilde{H}_\epsilon(U_0|Y_{-\Delta\tau})$ (unten).

das Verhalten, das auch bei der bedingten Entropie zwischen X und Y auftritt. Die Funktion $\tilde{H}_\epsilon(X_0|Y_{-\Delta\tau})$ ist in Abbildung 30 (oben) dargestellt. Die beiden Kurven in Abbildung 30 sind um die gleiche Zeitdifferenz wie $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$ und $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ (Abbildung 29) verschoben. Da diese Zeitdifferenz, wie im folgenden zu sehen sein wird, nicht willkürlich ist, kann aus diesem Verhalten gefolgert werden, daß die Informationsübertragung über die x -Komponente erfolgt.

Bislang wurden die Parameter k und c' aus Gleichung (7.3) festgehalten. Im weiteren wird nun untersucht, welchen Einfluß diese Parameter auf die Informationsübertragung von X auf U haben.

7.2.1 Der Einfluß der Phasenraumgeschwindigkeit k

In diesem Abschnitt soll die Verschiebung der Entropieminima in Abbildung 29 und 30 erklärt werden. Dies erfolgt durch den Vergleich mit der Phasendifferenz zwischen Anregung und resultierender Schwingung bei sinusförmig getriebenen Oszillatoren.

Zunächst wird der Parameter k , der die Zustandsraumgeschwindigkeit des Lorenzsystem bestimmt, betrachtet. Dieser entspricht der Frequenz ω im Fall eines linearen Antriebs von $\cos(\omega t)$. Für kleine k wird die Bewegung des Lorenzsystems verlangsamt, die verschiedenen Periodenlängen innerhalb des Systems werden folglich größer. In der u -Komponente stellen sich bei sehr kleinem k die Eigenschwingungen, die u.a. auch von den Anregungsamplituden abhängen, als hochfrequente Überlagerungen des Antriebs dar; auf der langsameren Zeitskala

des Lorenzsystems folgt die Dynamik in der u -Komponente der des antreibenden Systems. Die Entropiefunktionen $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$ und $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$ gleichen sich mit immer kleiner werdendem k zunehmend an. D.h. die bedingte Entropie und die beobachtete zeitliche Verschiebung der Extrema werden kleiner.

Bei größer werdendem k treten andere Effekte auf. Zunächst beginnt die bedingte Entropie zu wachsen, d.h. die Eigendynamik des Duffingsystems bekommt einen größeren Einfluß auf das Verhalten der u -Komponente, welcher durch Kenntnis des Lorenzsystems nicht vorherzusagen ist. Weiterhin wird die zeitliche Verschiebung größer; das Duffingsystem reagiert später auf die antreibende Kraft. Ein wesentlicher Punkt ist die stärkere Ausprägung des ersten lokalen Minimums von $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$. Dieses wandert, bei Nichtberücksichtigung der allgemeinen zeitlichen Verschiebung, auf einen Wert von ungefähr $0,4 s'/k$. Dies entspricht der halben Hauptperiodenlänge der z -Komponente des Lorenzsystems, deren Einfluß auf das gesamte Lorenzsystem bereits in Kapitel 5 gezeigt wurde.

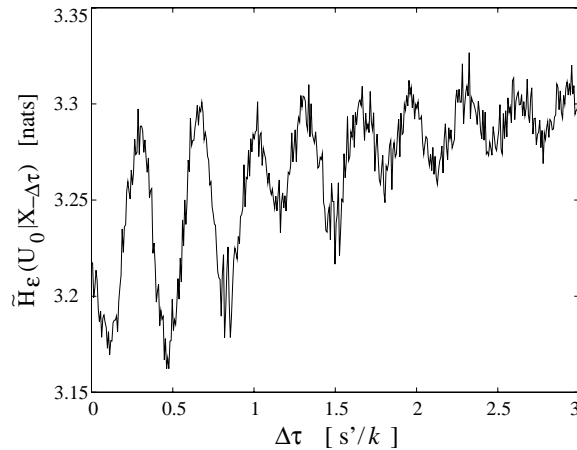


Abbildung 31: Bedingte Entropie $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$ bei $k = 0,35$.

Bei einem Wert von $k \simeq 0,35$ wird dieses Verhalten besonders auffällig (Abbildung 31). Die Entropiefunktion $\tilde{H}_\epsilon(U_0|X_{-\Delta\tau})$ oszilliert mit einer Periodenlänge von ca. $0,4 s'/k$, die zeitliche Verschiebung beträgt ungefähr $0,2 s'/k$. Im Bereich dieses k -Wertes treten Resonanzeffekte auf. Die u -Komponente schwingt mit der Hauptfrequenz des Lorenzsystems. Die Amplituden des Duffing-Oszillators werden dabei verstärkt; sie sind deutlich größer als bei kleinerem k . Wird k über den Resonanzbereich hinaus weiter erhöht, geht die Struktur der Entropiefunktion verloren. Eine weitere zeitliche Verschiebung kann zudem beobachtet werden,

weitere Minima sind aber nicht mehr vorhanden. Die schnelle Dynamik (der z -Komponente) des Lorenzsystems ist für die Dynamik des gekoppelten Systems, die nun durch die Eigenschwingungen des Duffingsystems geprägt ist, nicht mehr von entscheidender Bedeutung. Das Duffingsystem wird durch die langsame Dynamik des Lorenzsystems, also durch den Wechsel zwischen den charakteristischen Flügeln des Lorenzattraktors angeregt.

7.2.2 Einfluß der Kopplungsamplitude c'

Neben dem Einfluß von k muß auch derjenige der Kopplungsamplitude c' untersucht werden. Dazu werden für verschiedene Kopplungsamplituden die Kurven der bedingten Entropie berechnet (Abbildung 32) .

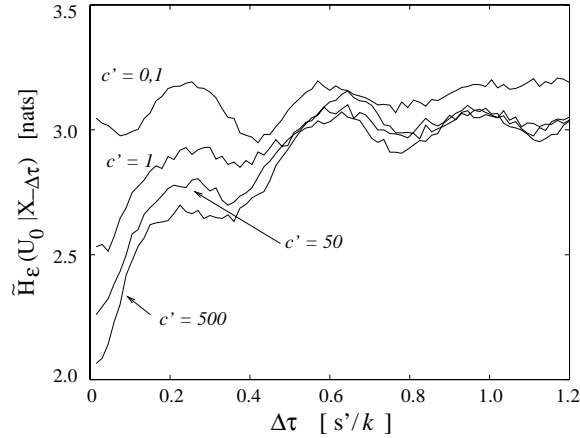


Abbildung 32: Bedingte Entropie $\tilde{H}_E(U_0|X_{-\Delta\tau})$ bei verschiedenen Kopplungsamplituden.

Die Kurven weisen das gleiche Verhalten auf wie bei einer Änderung von k . Das Verhalten bei großen Kopplungsamplituden entspricht dabei dem bei kleinen Phasenraumgeschwindigkeiten. Sowohl die zeitliche Verschiebung als auch die Struktur der Kurven zeigen deutliche Analogien zur oben vorgestellten Abhängigkeit von k . Auch der Übergang zu resonantem Verhalten ist zu erkennen: Eine Resonanz liegt in der Nähe von $c' = 0,1$ vor. Die Analogien zu den Entropieverläufen bei unterschiedlichen Phasenraumgeschwindigkeiten können durch die Abhängigkeit der Eigenfrequenzen des Duffing-Oszillators von der Anregungsamplitude erklärt werden, diese werden aufgrund des gegebenen harten Potentials mit wachsender Amplitude größer. Die Zeitverschiebung („Phasendifferenz“)

nimmt somit (bei fester Phasenraumgeschwindigkeit k) mit wachsendem c' ab.

Die Stärke des Einfluß des Lorenzsystem auf die u -Komponente des Duffing-Oszillators hängt also sowohl von der Kopplungsamplitude c' , als auch insbesondere von der Phasenraumgeschwindigkeit k ab.

7.2.3 Ergebnisse

Bei großem Einfluß des Antriebssystems X auf ein getriebenes System Y kann diese Abhängigkeit durch Berechnung der bedingten Entropien $\tilde{H}_\epsilon(Y_0|X_{-\Delta\tau})$ identifiziert werden: Diese zeigen eine ähnliche Struktur wie die Kurven $\tilde{H}_\epsilon(X_0|X_{-\Delta\tau})$; die Information des Antriebssystems fließt zu großen Teilen nach Y . Es ist sogar möglich, nähere Aussagen über die Art der Kopplung zu machen.

Weiterhin kann nichtlineares Resonanzverhalten erkannt werden. In diesem Fall werden zeitliche Abhängigkeiten innerhalb von X auf das getriebene System übertragen.

Wird der Einfluß der Dynamik des Antriebssystems jedoch geringer, ist die Identifizierung der Kopplung durch die Berechnung eindimensionaler Abhängigkeiten nicht mehr ohne weiteres möglich. Die Kurven $\tilde{H}_\epsilon(Y_0|X_{-\Delta\tau})$ enthalten keine charakteristischen Merkmale des Antriebssystems.

Die Methode kann somit einen allgemeinen informationstheoretischen Zugang zur Beschreibung der Synchronisation nichtlinearer dynamischer Systeme darstellen.

8 Bedingte Entropien in experimentellen Zeitreihen

8.1 Rückkopplungserscheinungen in Halbleiterlasern

Bereits in den 70er Jahren wurde festgestellt, daß dynamische Systeme mit zeitverzögerter Rückkopplung chaotisches Verhalten zeigen können. Für die Untersuchung an Lasersystemen ist besonders das sogenannte Ikeda-Szenarium [31] von Interesse.

8.1.1 Halbleiterlaser mit zwei externen Resonatoren

Bei einem Gleichstrom-getriebenen Halbleiterlaser mit zwei externen Resonatorspiegeln konnten Fischer et al. entsprechendes Verhalten experimentell nachweisen [32].

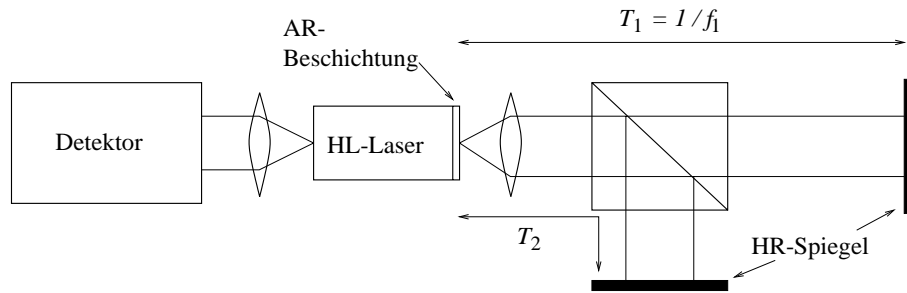


Abbildung 33: Schematischer experimenteller Aufbau (nach [32]).

In Abbildung 33 ist der dazugehörige Versuchsaufbau schematisch dargestellt. Das emittierte Licht durchläuft auf der rechten Seite einen Strahlteiler und wird dann von zwei Spiegeln zurück in den einseitig antireflektionsbeschichteten Laser reflektiert. Die Umlaufzeiten in den externen Resonatoren stehen im Verhältnis $\frac{T_1}{T_2} = \frac{2}{1}$. Die Intensität des linksseitig emittierten Lichts wird aufgezeichnet.

Mit diesem Aufbau konnten im Ikeda-Modell vorhergesagte Phänomene, insbesondere Übergänge von regulärem zu hochdimensionalen chaotischen Verhalten nachgewiesen werden. Ebenso wurde das Auftreten ungerader Harmonischer bzgl. der Frequenz $f_1 = 1/T$ beobachtet. Von Fischer und Heil wurden mit einem entsprechenden Versuchsaufbau für verschiedene Ströme Meßreihen der Lichtintensität aufgenommen. Ein Ausschnitt aus einer derartigen Zeitreihe ist in Abbildung

34 zu sehen.

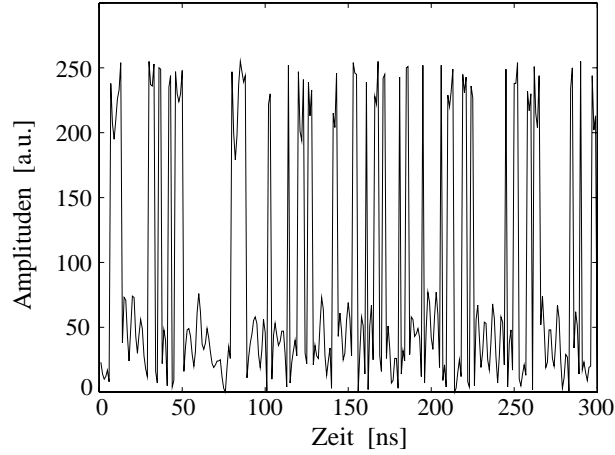


Abbildung 34: Ausschnitt aus einer Meßreihe der Lichtintensität bei einem Strom von 85 mA.

8.1.2 Zeitliche Abhängigkeiten

Die aufgenommenen Zeitreihen, die chaotisches Verhalten aufweisen, werden in dieser Arbeit mittels der Berechnung der modifizierten bedingten Entropie analysiert. Dazu wird wieder das in Kapitel 5 vorgestellte Verfahren verwendet, es werden also die eindimensionalen Abhängigkeiten $\tilde{H}_\epsilon(\Delta\tau)$ aus Gleichung (5.11) berechnet.

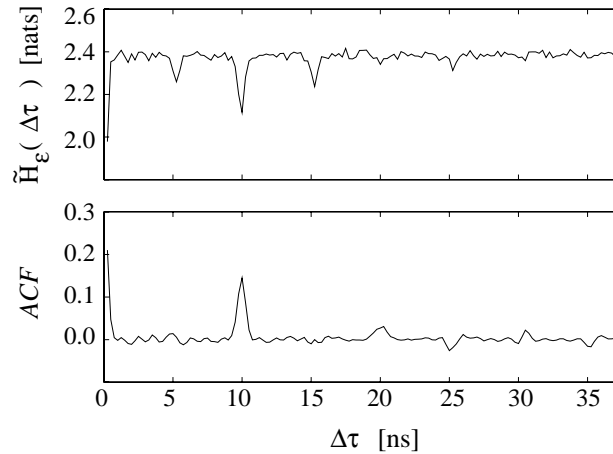


Abbildung 35: Bedingte Entropie (oben) und Autokorrelation (unten), berechnet aus der zu Abb.34 gehörenden Zeitreihe.

Es stellt sich heraus, daß entsprechend den gemessenen Umlaufzeiten $T_1 = 10$ ns und $T_2 = 5$ ns klare Minima im Entropieverlauf auftreten (Abb. 35, oben). Die Identifizierung dieser Umlaufzeiten ist mittels der Autokorrelationsfunktion nicht immer ohne weiteres möglich (Abb. 35, unten), obwohl sie bei den meisten Meßreihen für deren Bestimmung ausreicht. Doch auch in diesem Fall bietet die Berechnung der bedingten Entropie Vorteile. Aufgrund ihrer besseren Lokalisierbarkeit und zum Teil stärkeren Ausprägung eignen sich die Extrema der Entropiefunktion $\tilde{H}_\epsilon(\Delta\tau)$ besser zur Quantifizierung der bestehenden Abhängigkeiten.

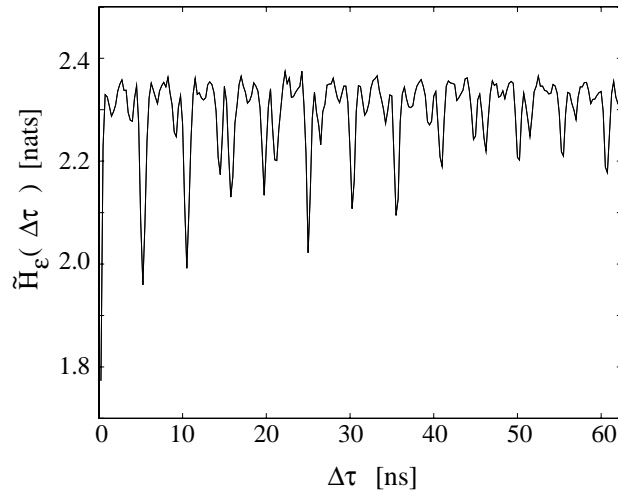


Abbildung 36: Bedingte Entropie $\tilde{H}_\epsilon(\Delta\tau)$ (aus einer bei einem Strom von 80 mA aufgenommenen Meßreihe berechnet).

Beim Auftreten der im Ikeda-Szenarium vorhergesagten dritten Harmonischen sind im Verlauf von $\tilde{H}_\epsilon(\Delta\tau)$ periodisch auftretende Minima zu erwarten. Die Amplituden dieser Minima der Entropiefunktion zeigen jedoch ein auf den ersten Blick irreguläres Verhalten (Abb. 36). Während bei kleinen Zeitdifferenzen das jeweils dritte Minimum, das erwartungsgemäß der ungefähren Resonatorumlaufzeit entspricht, den deutlich kleinsten Wert aufweist, erhält mit wachsendem $\Delta\tau$ das vorangehende Minima immer mehr Bedeutung. Bei weiter wachsender Zeitdifferenz wird dann das nächstfrühere Minmum ausgeprägter. Dies ist das typische Verhalten einer Schwebung. Eine Vermutung, daß dieses Verhalten ein Artefakt einer ungünstig gewählten Abtastrate sei, kann durch weitere Berechnungen nicht bestätigt werden.

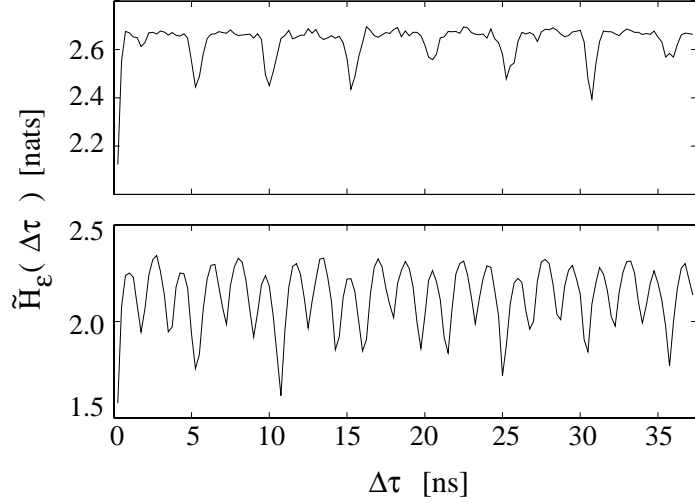


Abbildung 37: Bedingte Entropie $\tilde{H}_\epsilon(\Delta\tau)$ (aus zwei bei einem Strom von 75 mA aufgenommenen Meßreihen berechnet).

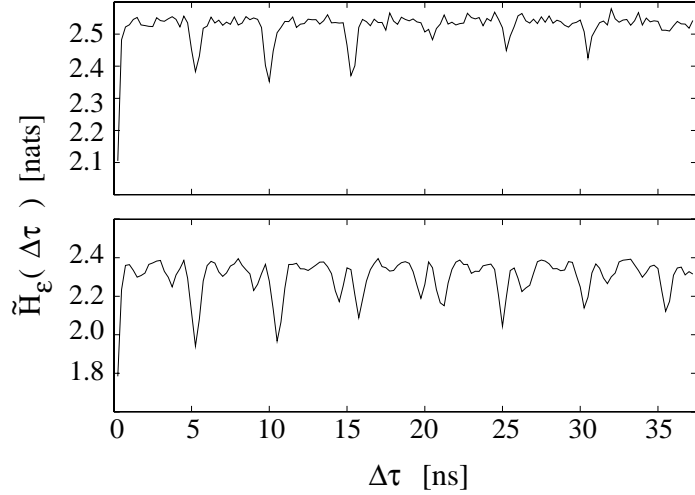


Abbildung 38: Bedingte Entropie $\tilde{H}_\epsilon(\Delta\tau)$ (aus zwei bei einem Strom von 80 mA aufgenommenen Meßreihen berechnet).

Zur Untersuchung dieses Phänomens werden weitere Entropiekurven aus Zeitreihen berechnet, die bei gleichen Nebenbedingungen aufgenommen wurden. Eine Zeitreihe weist eine Dynamik auf, bei der die Grundharmonische vorherrscht, die Dynamik der anderen wird durch die dritte Harmonische geprägt (Abb. 37). Dieses gleichzeitige Auftreten verschiedener Domänen wird durch das Ikeda-Modell vorhergesagt. Es stellt sich allerdings heraus, daß die Entropieminima dieser Zeitreihe nicht, wie vermutet, mit den jeweils dritten Minima der ersten

Kurve übereinstimmen. Das Verhältnis von der Grundfrequenz, welche weitgehend durch die Umlaufzeit im externen Resonator bestimmt ist ($\omega_1 \simeq 1/T_1$) zu der vorherrschenden Frequenz der schnelleren Dynamik ω_2 ist nicht $\frac{3}{1}$, sondern $\frac{\omega_1}{\omega_2} \simeq \frac{14}{3}$.

Sowohl die schnellere als auch die langsamere (Grund-) Dynamik haben einen Einfluß auf die emittierte Lichtintensität. Je besser die Vielfachen der Periodenlängen übereinstimmen, desto mehr Information über die entsprechenden zukünftigen Zustände kann gewonnen werden. Die bedingte Entropie der „dritten Harmonischen“ ist also dann besonders niedrig, wenn auch bei der Entropie der Grundharmonischen ein Minima auftritt (Abb. 37 und 38). Da dies auch für den umgekehrten Fall gilt, kann aus dem Verschwinden der Minima der Entropiefunktion in Abbildung 35 auch auf die Existenz „höherer Harmonischer“ geschlossen werden. Das Auftreten der beobachteten Verstimmungen ist ein Unterschied zum Ikeda-Szenarium, zu welchem das Systemverhalten qualitativ viele Ähnlichkeiten aufweist.

Aufgrund der Größe der Frequenzverschiebung ist der Effekt auch nicht direkt auf eine Abweichungen der Umlaufzeiten T_1 und T_2 vom eingestellten Verhältnis $\frac{2}{1}$ zu erklären. Die Berechnung der modifizierten bedingten Entropie bei den vorhandenen Meßreihen deutet auf eine Abhängigkeit der Art und Größe der Verstimmung vom antreibenden Strom hin. Mit wachsender Stromstärke scheint das Frequenzverhältnis $\frac{\omega_1}{\omega_2}$ größer zu werden. Zur Verifizierung dieser Annahme reicht das vorhandene Datenmaterial jedoch nicht aus. Weitere Messungen könnten darüber Aufschluß geben.

8.2 Quasiperiodisch getriebener, nichtlinearer Schwingkreis

Um den Einfluß verschiedener Frequenzen auf die Entropiefunktion $\tilde{H}_\epsilon(\Delta\tau)$ weiter zu untersuchen, wird das Verhalten eines quasiperiodisch getriebenen, nichtlinearen elektrischen Schwingkreises vermessen. Dieser besteht aus einem Widerstand $R = 200 \, \Omega$, einer Induktivität $L = 6,4 \, \text{mH}$ und einer spannungsabhängigen Kapazitätsdiode $C(U)$ (Typ 1N4007) (Abbildung 39). Dieses System kann angenähert durch einen nichtlinearen Oszillator mit Toda-Potential $\Phi(x) = \exp(x) - x - 1$ beschrieben werden.

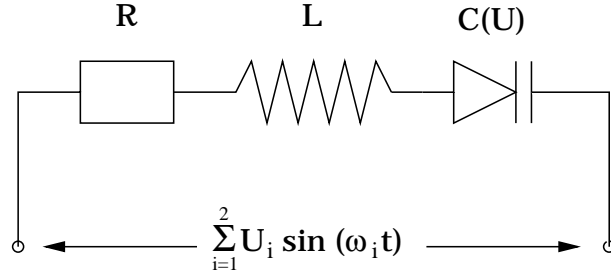


Abbildung 39: Quasiperiodisch getriebener, nichtlinearer elektrischer Schwingkreis.

Mittels zweier Frequenzgeneratoren werden zwei Sinuswellen mit Amplitude U_i und Frequenz ω_i erzeugt und anschließend addiert. Mit dieser überlagerten Schwingung wird das System getrieben. Am Widerstand des Schwingkreises wird der resultierende Strom gemessen. Die Eingangsfrequenzen werden derart gewählt, daß sie in einem annähernd rationalen Verhältnis zueinander stehen. Zwei Zeitreihen des Systems werden untersucht. Die erste (Z1) wurde bei einem Spannungsverhältnis von $\frac{U_1}{U_2} \simeq \frac{1,1 \text{ V}}{1,9 \text{ V}}$ und einem Frequenzverhältnis von $\frac{\omega_1}{\omega_2} \simeq \frac{99,9 \text{ kHz}}{140 \text{ kHz}} \simeq \frac{5}{7}$ aufgenommen, die zweite (Z2) bei $\frac{U_1}{U_2} \simeq \frac{2,0 \text{ V}}{1,9 \text{ V}}$ und $\frac{\omega_1}{\omega_2} \simeq \frac{159,9 \text{ kHz}}{140 \text{ kHz}} \simeq \frac{8}{7}$. Diese Zeitreihen weisen chaotisches Verhalten auf (Abbildung 40).

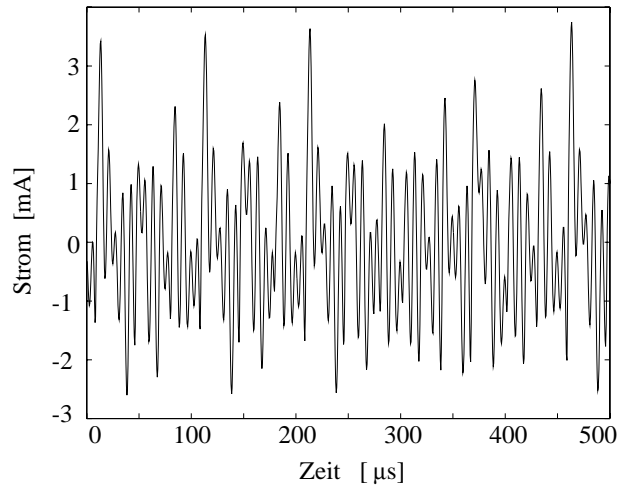


Abbildung 40: Ausschnitt aus der Zeitreihe Z1, die bei einem Frequenzverhältnis von $\frac{\omega_1}{\omega_2} \simeq \frac{5}{7}$ aufgenommen wurde.

Die Berechnung der modifizierten bedingten Entropie aus diesen beiden

Zeitreihen liefert ein ähnliches Ergebnis wie bei der Untersuchung der Intensitätszeitreihen des rückgekoppelten Halbleiterlasers. Für beide Zeitreihen kann aus der periodischen Abfolge der Minima der Entropiefunktion $\tilde{H}_\epsilon(\Delta\tau)$ wiederum jeweils eine dominierende Frequenz ermittelt werden. Ebenso werden variierende Amplituden dieser Minima beobachtet (Abbildungen 41 und 42).

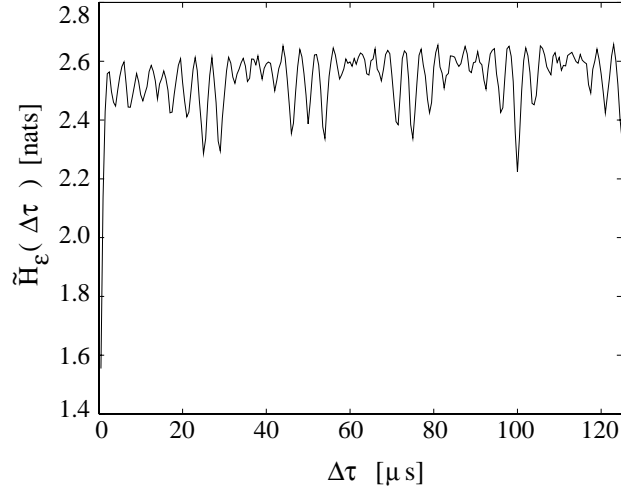


Abbildung 41: Bedingte Entropie $\tilde{H}_\epsilon(\Delta\tau)$ bzgl. der Zeitreihe Z1.

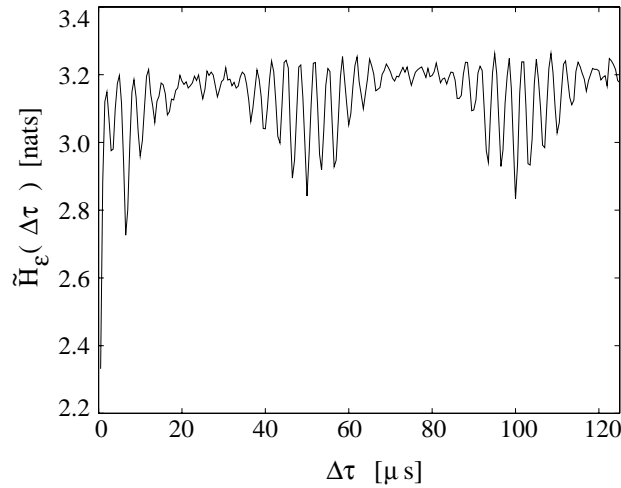


Abbildung 42: Bedingte Entropie $\tilde{H}_\epsilon(\Delta\tau)$ bzgl. der Zeitreihe Z2.

In beiden Fällen treten die Minima in Abständen von $\Delta\tau \simeq 3,6 \mu s$ auf. Dieser Wert entspricht der halben Periodenlänge der durch U_2 bestimmten Anregungsschwingung. In Abbildung 41 ist zu erkennen, daß die einzelnen Minima

in Abständen von $\Delta\tau \simeq 25 \mu\text{s}$ besonders stark ausgeprägt erscheinen. Diese Zeitdifferenz entspricht der Differenzfrequenz $\Delta\omega = |\tilde{\omega}_1 - \tilde{\omega}_2| \simeq 40 \text{ kHz}$. Bei der Entropiefunktion von Zeitreihe Z2 ist $\Delta\tau \simeq 50 \mu\text{s}$, also $\Delta\omega \simeq 20 \text{ kHz}$. Auch dieser Wert stimmt erwartungsgemäß mit der Frequenzdifferenz der gewählten Anregungsschwingungen überein.

Zusammenfassend ist zu bemerken, daß das Auftreten von Schwingungen unterschiedlicher Frequenz innerhalb eines chaotischen Systems zu Informationsverlusten bzgl. fester Vorhersagezeiten führen kann, welche durch die jeweiligen Frequenzen bestimmt sind. Dies drückt sich durch eine Amplitudenmodulation in der modifizierten bedingten Entropie aus. Dabei zeigt die Entropiekurve $\tilde{H}_\epsilon(\Delta\tau)$ periodisch auftretende Minima, die mit der dominierenden Frequenz des Systems korrespondieren. Die Größe dieser Minima hängt von den Frequenzdifferenzen zu den anderen Schwingungen ab. In chaotischen Zeitreihen, in denen derartige Periodizitäten nicht mehr mittels linearer Standardverfahren eindeutig bestimmt werden können, ermöglicht die Berechnung der modifizierten bedingten Entropie somit die Identifizierung verborgener Periodizitäten.

Damit erhärten diese Ergebnisse die Vermutung, daß eine Verstimmung zwischen den dominierenden Frequenzen der Zeitreihen des Lasersystems vorliegt.

9 Datengestützte Modellbildung

In Kapitel 4 wurde gezeigt, wie mittels der Berechnung der modifizierten bedingten Entropie die Meßgrößen X_i ermittelt werden können, welche am besten zur Vorhersage einer anderen Größe Y geeignet sind. Wenn die bedingte Entropie $\tilde{H}_\epsilon(Y|(X_1, \dots, X_N))$ verschwindet, existiert eine stetige Funktion, durch die Y fast sicher aus den X_i hervorgeht. Allerdings liegt damit noch keine Aussage über die spezielle Form dieser Funktion vor. In diesem Kapitel werden nun anhand von Beispielen Methoden vorgestellt, mit denen entsprechende mathematische Modelle gefunden werden können.

9.1 Parameterermittlung

Im Gegensatz zu physikalischen Modellen werden durch mathematische Modelle keine Ursache-Wirkung-Beziehungen beschrieben, sondern diejenigen Funktionen gesucht, welche gegebene Daten am besten beschreiben können. Damit ist die mathematische Modellierung eine Form der Fehlerminimierung.

In der Regel sind Eingangsdaten $\mathbf{x}_t \in \mathbb{R}^d$ und Ausgangsdaten y_t , im vorliegenden Fall skalare Werte, gegeben. Es wird dann eine Menge von Basisfunktionen $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$, $f_i(\mathbf{a}_{i,j}, \mathbf{x}_t) \mapsto y_t$ vorgegeben. Jede dieser Basisfunktionen hat einen oder mehrere freie Parameter $\mathbf{a}_{i,j}$. Bei der Modellbildung werden dann diejenigen Parameter gesucht, mit denen das Modell die Ausgangsgröße Y am besten vorhersagt.

Werden alle vorhandenen Daten verwendet, so können diese Parameter durch die Berechnung des χ^2 -Fehlers bestimmt werden [13]. Dieser ist durch die Summe der quadrierten Abstände zwischen den Ausgangsdaten und den vorhergesagten Punkten gegeben

$$\chi^2 = \sum_{t=1}^{N_T} \frac{(y_t - f(\mathbf{a}_{i,j}, \mathbf{x}_t))^2}{\sigma_t}. \quad (9.1)$$

Dabei stellt f eine Kombination der Basisfunktionen f_i dar, und σ_t ist in diesem Fall die Unsicherheit, welche über die Meßgrößen zur Zeit t besteht. Können keine Aussagen über diese Unsicherheit getroffen werden, so wird $\sigma_t = 1$ gesetzt.

Die besten Parameter können nun mittels des „simulated annealing“ [13] ermittelt werden. Dieses Verfahren ist an thermodynamische Prozesse bei der

Abkühlung eines Systems angelehnt. Dabei kommt es zu Phasenübergängen; das System nimmt Zustände niedriger Energie an. Erfolgt die Abkühlung schnell, kann es zu kristallinen Strukturen kommen, die nicht unbedingt den Zustand der absolut niedrigsten Energie darstellen, welcher bei langsamer Abkühlung eingenommen wird. Bei der Parameterbestimmung entspricht der χ^2 -Fehler der Energiefunktion. Die thermodynamischen Prozesse werden durch zufällige Transformationen eines Simplexes im Parameterraum simuliert. Dieser „hat das Bestreben“, Zustände möglichst kleiner Energie, also kleiner quadratischer Fehler einzunehmen. Bei einem hohen, einstellbaren „Temperatur-Parameter“ erstrecken sich die Bewegungen des Simplex über große Bereiche des Parameter-raums. Abkühlung, d.h. Verkleinerung dieses Parameters sorgt dafür, daß der Simplex auf immer kleinerem Gebiet in Richtung eines Energieminimums strebt, das bei hinreichender Abkühlung erreicht wird. Erfolgt diese schnell, ist es allerdings möglich, daß er lediglich ein lokales Minimum erreicht. In diesem Fall ist es angebracht, das Verfahren bei unterschiedlichen Startbedingungen mehrfach zu wiederholen.

Die Bestimmung optimaler Modellparameter vereinfacht sich extrem, wenn die Basisfunktionen linear in den Parametern sind, dann ist

$$f(\mathbf{a}_{i,j}, \mathbf{x}_t) = \sum_{i=1}^M a_i f_i(\mathbf{x}_t).$$

Durch die N_T Funktionswerte zu den M Basisfunktionen ist eine $(N_T \times M)$ -Matrix D , die sogenannte *Design-Matrix* gegeben. Damit beschränkt sich die Ermittlung der Parameter a_i auf die Lösung des linearen Gleichungssystems

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_T} \end{pmatrix} = \begin{pmatrix} f_1(\mathbf{x}_1) & f_2(\mathbf{x}_1) & \cdots & f_M(\mathbf{x}_1) \\ f_1(\mathbf{x}_2) & f_2(\mathbf{x}_2) & \cdots & f_M(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(\mathbf{x}_{N_T}) & f_2(\mathbf{x}_{N_T}) & \cdots & f_M(\mathbf{x}_{N_T}) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix} \quad (9.2)$$

Ist das Gleichungssystem überbestimmt, so liefert, wie in Kapitel 4 bereits erwähnt, eine Singulärwertzerlegung das im Sinne der kleinsten Fehlerquadrate beste Ergebnis [13]. Dies ist genau das erwünschte Resultat.

9.2 Termauswahlverfahren

Bislang wurden alle vorhandenen Datenpunkte zur Modellbildung verwendet. Die Parameterbestimmung mittels Berechnung des χ^2 -Fehlers liefert das für den entsprechenden Datensatz beste Modell. Das bedeutet jedoch nicht, daß dieses Modell auch für andere Daten desselben Systems optimal angepaßt ist. So ist es z.B. denkbar, daß stochastische Störungen in die Parameterschätzung einfließen. Diese sind aber nicht für das Modell relevant und würden bei anderen Daten des gleichen Systems für größere Fehler sorgen.

Um dies zu vermeiden, können die vorhandenen Daten in einen Trainings- und einen Testdatensatz aufgeteilt werden. Das Modell wird anhand des Trainingsdatensatzes erstellt und die Abweichung des Modells von den Realdaten anhand des Testdatensatzes berechnet [33].

Eine extreme Variante der Aufteilung des Datensatzes in Trainings- und Testdaten kann erfolgen, indem die Testdaten nur aus einem einzigen Datenpunkt bestehend gewählt wird. Dabei werden für alle N_T Datenpunkte Modelle erstellt und dann ein mittlerer Modellfehler

$$\sigma_{LOO}^2 = \frac{1}{N_T} \sum_{t=1}^{N_T} (y_t - f_t(\mathbf{x}_t))^2 \quad (9.3)$$

berechnet, wobei die Modelle f_t aus den jeweils verbleibenden $N_T - 1$ Datenpunkten (\mathbf{x}_s, y_s) , $s \neq t$ mittels der Berechnung des χ^2 -Fehlers erstellt werden. Bei Modellen, welche linear in den Parametern sind, hat dieser „Leave-One-Out“-Fehler den Vorteil, daß er analytisch berechnet werden kann [33]. Mit der Projektionsmatrix

$$B = I_{N_T} - D(D^\top D)^{-1} D^\top,$$

wobei I_{N_T} die N_T -dimensionale Einheitsmatrix und D^\top die Transponierte der Designmatrix darstellt, ist

$$\sigma_{LOO}^2 = \frac{1}{N_T} \cdot \vec{y}^\top B (\text{diag}(B))^{-2} B \vec{y}. \quad (9.4)$$

Der Vektor \vec{y} besteht aus den N_T y -Werten des gesamten Datensatzes. Mittels Gleichung (9.4) ist eine schnelle Berechnung des Modellfehlers σ_{LOO}^2 möglich.

Während der χ^2 -Fehler mit wachsender Anzahl M an Basisfunktionen monoton kleiner wird, weist der σ_{LOO}^2 -Fehler bei einer optimalen Menge von Basisfunktionen ein absolutes Minimum auf. Dieser Satz von Basisfunktionen kann

iterativ ermittelt werden. Dazu kann mit der Berechnung des σ_{LOO}^2 -Fehlers begonnen werden, wenn sämtliche vorhandenen Basisfunktionen gegeben sind, um dann sukzessiv diejenigen Funktionen aus der Funktionenmenge zu entfernen, die den σ_{LOO}^2 -Fehler am stärksten erhöhen. Umgekehrt kann auch von einer leeren Funktionenmenge ausgegangen werden. Dann werden solange diejenigen Basisfunktionen, welche den σ_{LOO}^2 -Fehler am stärksten minimieren, hinzugenommen, bis das Minimum erreicht wird. Diese zweite Methode, die in dieser Arbeit verwendet wird, heißt „Forward Selection“⁷. Die endgültige Parameterermittlung erfolgt schließlich durch die Anwendung der im vorherigen Abschnitt beschriebenen Methode der kleinsten Fehlerquadrate auf den ausgewählten Satz von Basisfunktionen.

Als Basisfunktionen sind d -dimensionale Polynome p -ter Ordnung

$$g_i(\mathbf{x}) = x_1^{j_1} \cdot x_2^{j_2} \cdots x_d^{j_d},$$

mit $\sum_k j_k \leq p$, sehr gut geeignet. Die Polynomialmodelle

$$f(\mathbf{a}_{i,j}, \mathbf{x}_t) = \sum_{i=1}^M a_i g_i(\mathbf{x}_t)$$

stellen eine natürliche Erweiterung linearer Modelle dar. Außerdem lassen sich analytische Funktionen durch Potenzreihenentwicklung in dieser Form darstellen.

9.2.1 Analyse von Turbogeneratoren

Als Beispiel für die Forward Selection wird nun die Modellbildung von Schwingungszuständen von Turbomaschinen gezeigt. Beobachtungen in Kraftwerken zeigen, daß einige Turbogeneratoren jeden Abend zur Vollastphase eine starke Überhöhung der Schwingungsamplituden aufweisen (Abbildung 43).

Als Eingangsgrößen wurden 35 verschiedene Betriebsparameter X_i wie Erregerstrom, Abdampfdrücke und -temperatur, sowie die Wirkleistung gemessen. Die gemessenen Ausgangsgrößen Y sind die maximalen Schwingungsamplituden der Turbogeneratoren. Unter der Annahme, daß diese weitgehend durch Erregerstrom X_2 und Wirkleistung X_1 beschrieben werden können, wurden mittels Forward Selection aus diesen Größen Polynomialmodelle mit $p = 2$ gebildet.

⁷Um eine mögliche bei der Iteration auftretenden Fehlausewahl der relevanten Terme zu vermeiden, können auch beide Vorgehensweise miteinander verbunden werden.

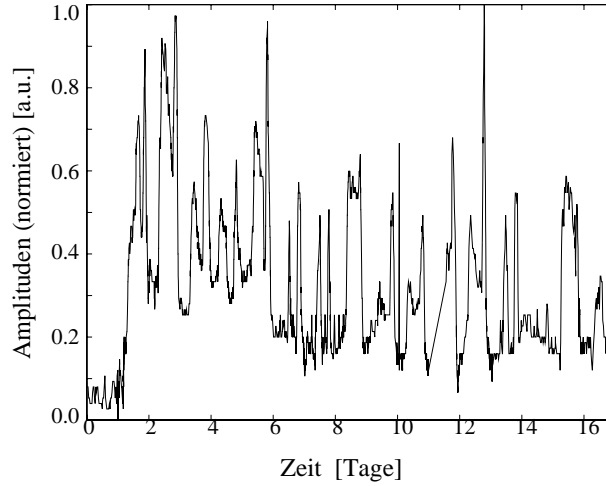


Abbildung 43: Zeitlicher Verlauf der maximalen Schwingungsamplituden eines Turbogenerators.

Während die Schwingungsamplituden ohne Überhöhung am besten durch ein Modell der Form

$$Y = a_0 + a_1 X_1 + a_2 X_2^2 \quad (9.5)$$

beschrieben werden können, ist für die Zeitreihen mit überhöhten Amplituden ein Modell der Form

$$Y = a_0 + a_1 X_2 + a_2 X_2^2 \quad (9.6)$$

am besten geeignet. In diesem Fall ist also die Wirkleistung X_1 für das Modell ohne Bedeutung [34].

9.3 Bedingte Entropie und Modellbildung

9.3.1 Zusammenhänge zwischen den Meßgrößen

Dieses Ergebnis soll unter dem Gesichtspunkt der bedingten Entropie untersucht werden. Dazu werden die modifizierten bedingten Entropien der Schwingungsamplituden Y für jeden gegebenen Betriebsparameter X_i für den gesamten Schwingungsverlauf berechnet (Abbildung 44).

Es zeigt sich, daß der Erregerstrom X_2 die geringste und die Wirkleistung X_1 die zweitkleinste bedingte Entropie ergibt. Dies scheint die Annahme, daß diese beiden Größen zur Modellbildung am besten geeignet sind, zu bestätigen. Um

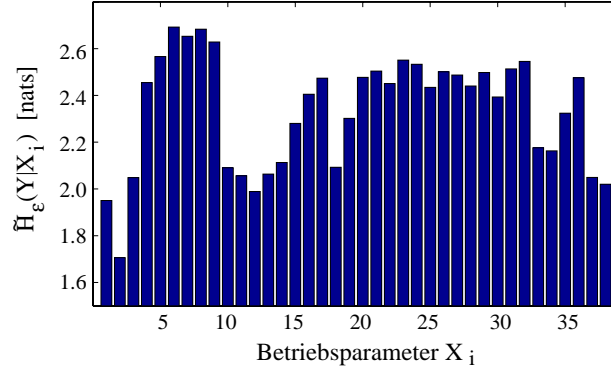


Abbildung 44: Bedingte Entropien $\tilde{H}_\epsilon(Y|X_i)$ der Schwingungsamplituden bei verschiedenen gegebenen Betriebsparametern X_i .

dies zu überprüfen wird im zweiten Schritt X_2 festgehalten und $\tilde{H}_\epsilon(Y|(X_2, X_i))$ berechnet.

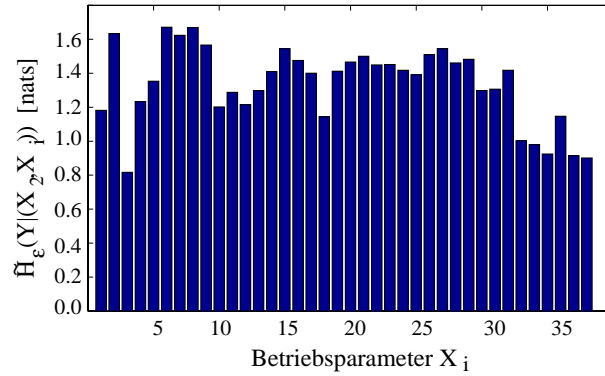


Abbildung 45: Bedingte Entropien $\tilde{H}_\epsilon(Y|(X_2, X_i))$ der Schwingungsamplituden bei gegebenem Erregerstrom X_2 und anderen Betriebsparametern X_i .

In Abbildung 45 ist zu erkennen, daß bei festgehaltenem Erregerstrom die Wirkleistung nicht mehr die bedingte Entropie minimiert, sondern die Abdampftemperatur X_3 . Die in der Wirkleistung über die Schwingungsamplituden enthaltene Information ist zu einem großen Teil bereits durch den Erregerstrom gegeben. Dies ist die Erklärung dafür, daß die Wirkleistung in die Modellbildung (9.6) nicht einfließt. Wird jedoch die Abdampftemperatur X_3 hinzugenommen, ergibt die Forward Selection ein Modell der Form

$$Y = a_0 + a_1 X_2 + a_2 X_2^2 + a_3 X_3. \quad (9.7)$$

Die Hinzunahme des linearen Terms $a_3 X_3$ reduziert den Modellfehler um ca. 10%.

9.3.2 Vorgehensweise bei der Modellbildung

Das aufgeführte Beispiel weist einen geeigneten Weg zur datengetriebenen Modellbildung auf, bei der keine weiteren Kenntnisse über das zu beschreibende System einfließen müssen.

- Im ersten Schritt werden iterativ durch Berechnung der modifizierten bedingten Entropie diejenigen Größen ermittelt, welche zur Vorhersage der zu modellierenden Meßgröße am besten geeignet sind.
- Diese Größen werden dann für die Forward Selection verwendet. Durch diese werden die für das Modell relevanten Terme bestimmt. Gleichzeitig kann eine Parameterabschätzung mittels Singulärwertzerlegung erfolgen.

Das vorgestellte Verfahren kann für verschiedene Situationen weiter modifiziert werden.

Es ist möglich, daß sich die Abhängigkeiten der Ausgangsgrößen von den Eingangsgrößen mit den Zuständen der X_i ändern. Um derartige nichtstationäre Systeme modellieren zu können, kann eine Clusterung im Raum der Eingangsgrößen vorgenommen werden. In dieser Arbeit wurde ein Fuzzy-Clustering (FCM-Clustering [35]) verwendet, das ermöglicht, die Clustergrenzen fließend zu wählen. Für die Entropieberechnung sind harte Cluster, bei denen jeder Punkt genau einem Cluster zugeordnet ist, zu verwenden, damit eindeutige Zusammenhänge zwischen den Meßgrößen bestimmt werden können. Bei der Modellbildung ist es aber häufig sinnvoll, weiche Clustergrenzen zu wählen. Dabei werden die Punkte allen Clustern mit verschiedenen Gewichten zugeordnet⁸. Somit können Übergänge zwischen den einzelnen Clustern modelliert werden.

Bei kontinuierlichen dynamischen Systemen ist ein Modell in Form eines Differentialgleichungssystems gemäß (2.3) wünschenswert. In diesem Fall sind die Ausgangsgrößen Y_i die Ableitungen an den jeweiligen Punkten der Zeitreihe. Da diese in der Regel jedoch nicht explizit vorliegen, müssen sie aus der Zeitreihe selbst berechnet werden. Im einfachsten Fall können dazu die Differenzen aufeinander folgender Punkte verwendet werden. Dabei treten aber, insbesondere bei

⁸Harte Cluster ergeben sich als Spezialfall, bei dem jeweils ein Cluster das Gewicht 1, die anderen das Gewicht 0 erhalten. Im anderen Extrem besitzen alle Cluster das gleiche Gewicht.

Vorhandensein von Rauschen, extreme Fehler auf. Erheblich bessere Ergebnisse liefert das lokale Anfitten der Meßreihen mittels differenzierbarer Funktionen, beispielsweise durch Savitzky-Golay-Filter [13], und deren anschließende analytische Differentiation. Mit diesen Ergebnissen können sowohl die Berechnung der modifizierten bedingten Entropie, als auch die Modellbildung durchgeführt werden. Allerdings treten auch bei dieser Abschätzung der Ableitungen Fehler auf, so daß es häufig sinnvoll ist, zur Parameterabschätzung ein weiteres Optimierungsverfahren wie z.B. das simulated Annealing zu verwenden. Dabei wird die gefundene Differentialgleichung bei verschiedenen Startparametern integriert, wobei die Energiefunktion durch die quadratische Abweichung von den Meßdaten bestimmt ist.

Das vorgestellte Verfahren ermöglicht es, aus mehrdimensionalen Daten ohne tiefere Kenntnis des Systems Modelle zu erstellen. Es muß jedoch darauf hingewiesen werden, daß es sich dabei um mathematische Modelle handelt, also Funktionen, welche Meßdaten mit dem kleinstmöglichen Fehler vorhersagen. Weder durch die Modellbildung noch durch die Berechnung der bedingten Entropie können direkte Aussagen über Ursache-Wirkung-Verhältnisse gemacht werden. Die in [36] aus den Ergebnissen der Modellbildung (9.6) gezogene Schlußfolgerung, daß der Erregerstrom keinen Einfluß auf die Wirkleistung hat, ist nicht zwingend, zumal dieser viel Information über die Leistung enthält, also eine kleine bedingte Entropie liefert.

Die hier vorgestellte Vorgehensweise kann nur Hinweise auf die physikalischen Zusammenhänge geben. Sie stellt keinen Automatismus dar: bei der Interpretation der Ergebnisse muß in jedem Schritt ihre physikalische Relevanz überprüft werden.

9.4 Echtzeitanalyse

Teil der Arbeit ist, die vorgestellten Methoden auf ihre Einsetzbarkeit für Echtzeitanalysen zu überprüfen [37]. Im Rahmen dieser Arbeit wurde daher ein komplexes Analysewerkzeug entwickelt (Anhang C). Die einzelnen Analysemethoden werden dabei als eigenständige Programme implementiert. Sie können dabei wie in Abbildung 46 dargestellt miteinander verknüpft werden. Dieser Aufbau be-

dingt, daß der langsamste Prozeß die Gesamtgeschwindigkeit der Berechnungen in jedem Schritt vorgibt. Es kommt also darauf an, möglichst schnelle Algorithmen zur Berechnung der bedingten Entropie und der Modellbildung zu finden. Je länger die Zeitreihe ist, desto mehr Zeit wird für die einzelnen Analysen benötigt. Daher wird der Datenstrom zunächst in Teilsequenzen von einigen hundert bis wenigen tausend Datenpunkten aufgeteilt.

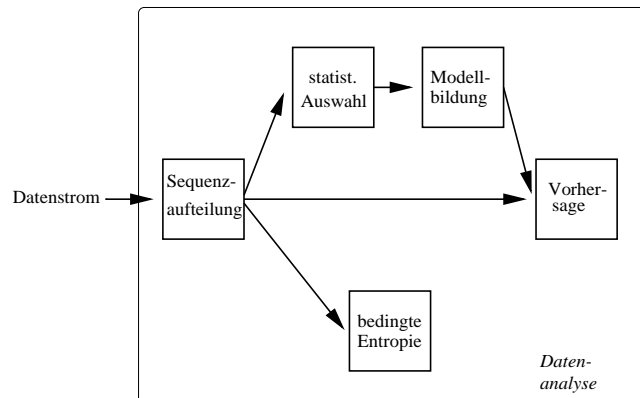


Abbildung 46: Verschaltung verschiedener Prozesse bei der Echtzeitanalyse.

Bei der Berechnung der bedingten Entropie ist vor allem die Suche nach den nächsten Nachbarn eines Referenzpunktes zeitaufwendig, insbesondere in hochdimensionalen Systemen. Da die Nachbarsuche für sehr viele Algorithmen in der Analyse nichtlinearer Zeitreihen benötigt wird, wurden bereits eine Reihe von Suchverfahren entwickelt, z.B. die Verwendung d -dimensionaler Suchbäume [38]. Da bei der beschränkten Datenzahl, die für die Echtzeitanalyse verwendet werden, lediglich ein- bis zweidimensionale Abhängigkeiten untersucht werden können, sind derartige Verfahren nicht unbedingt notwendig. Dennoch ist eine Vorsortierung der Daten zweckmäßig, damit nicht wiederholt die Abstände zwischen allen Datenpunkten berechnet werden müssen. Im eindimensionalen Fall kann dies einfach durch das Sortieren der Daten nach ihrer Größe (beispielsweise mittels des Quicksort-Algorithmus [39]) geschehen. Bei der Suche nach 2-dimensionalen Nachbarpunkten ist eine Aufteilung des gesamten besetzten Zustandsraums in Quadrate der Seitenlänge ϵ sinnvoll. Dadurch beschränkt sich die Abstandsrechnung auf die Punkte, die in dem Quadrat des Referenzpunkts liegen, und diejenigen der maximal acht Nachbarquadrate [38]. Eine weitere Reduzierung der Rechenzeit kann erfolgen, indem nicht mehr alle Datenpunkte zur Berechnung

des Mittelwertes verwendet werden, sondern lediglich eine sinnvolle, statistische Auswahl.

Eine derartige Auswahl ist auch für die Modellbildung von Nutzen. Trotz dieser Vorauswahl und der schnellen analytischen Berechnung des σ_{LOO}^2 -Fehlers dauert die Berechnung eines Modells ungefähr 1 s ⁽⁹⁾. Dies ist für eine Vielzahl von technischen und physikalischen Anwendungen zu langsam. Der Einsatz adaptiver Methoden könnte die Modellierung beschleunigen. Das in [40] vorgeschlagene adaptive Verfahren hat jedoch den Nachteil, daß es für jeden Modellbildungsschritt Einschwingzeiten von einigen hundert Zeitpunkten benötigt. In diesen Zeiten ist keine Aussage über die exakte Modellstruktur möglich. Bei der Prozeßüberwachung ist ein derartiges Verhalten ausgesprochen problematisch. In diesem Fall ist es günstiger, ein festes Modell zu verwenden und die dabei auftretenden Modellfehler zu berechnen. Nichtstationäre Systeme werden dann durch die Verwendung von Fuzzy-Clustern, wie oben beschrieben, modelliert. Um schnellere Analysemethoden nicht zu verzögern, erfolgt dann die Modellbildung losgelöst vom restlichen Analyseprozeß. D.h. es werden solange „alte“ Modelle verwendet, bis ein neues erstellt worden ist; das Modell wird nicht den aktuellen Daten angepaßt. Da die Modelle im nichtstationären Fall jedoch von der Lage der Eingangsgrößen im Zustandsraum abhängen, müssen sie auch dem aktuellen Zustandsraumbereich der Eingangsdaten entsprechen. Die Zuteilung der Teilmodelle zu bestimmten Clustern, muß somit weiterhin für den gesamten Raum der Eingangsgrößen gültig sein. Es ist demzufolge notwendig, die zuvor verwendeten Daten zu speichern, um eine dem Gesamtsystem entsprechende Verteilung der Eingangsdaten zu ermöglichen. Die permanente Speicherung aller Daten würde jedoch die Speicherkapazität des Computers überschreiten. Aus diesem Grund erfolgt eine statistische Auswahl. Sobald eine zuvor angegebene Speicherkapazität erreicht ist, werden die gespeicherten Daten zu einem großen Teil beibehalten. Der zufällig ausgewählte Rest wird durch einen ebenfalls zufällig ausgewählten Anteil aktueller Daten ersetzt. Auf diese Weise wird gewährleistet, daß die gespeicherten Daten die gleiche Verteilung aufweisen wie die Gesamtdaten. Zur Prozeßüberwachung kann damit einerseits der Modellfehler, andererseits (auf einer langsameren

⁹Auf einer Sun-Ultra10-Workstation, bei Verwendung von 500 Datenpunkten, einem Polynommodell dritter Ordnung und vier Eingangsgrößen.

Zeitskala) die Änderung des Modells verwendet werden.

10 Andere Methoden zur Bestimmung von Abhängigkeiten innerhalb von Zeitreihen und der Einfluß von Rauschen auf die Berechnung der bedingten Entropie

In diesem Kapitel soll auf die Vorteile und Grenzen des in dieser Arbeit vorgestellten Verfahrens eingegangen werden. Zunächst wird dabei die Anwendbarkeit der vorgestellten Methode bei verrauschten Daten untersucht. Im zweiten Abschnitt soll sie dann mit anderen Verfahren verglichen werden.

10.1 Der Einfluß von Meßrauschen

Die hier vorgestellte Methode zur Berechnung einer modifizierten bedingten Entropie hat sich als gut geeignet zur Analyse von Zeitreihen nichtlinearer Systeme erwiesen. In diesem Abschnitt wird nun der Einfluß von Meßrauschen auf die Analyseergebnisse untersucht. Dies soll anhand der in Kapitel 6 vorgestellten Methode zur Ermittlung optimaler Einbettungsparameter am Beispiel des Lorenzsystems geschehen. Dies ist ein kompliziertes Problem, da sowohl Nichtlinearitäten auftreten, als auch das System zu Beginn des Iterationsverfahrens in Bezug auf die Anzahl unabhängiger Variablen nicht vollständig bestimmt ist.

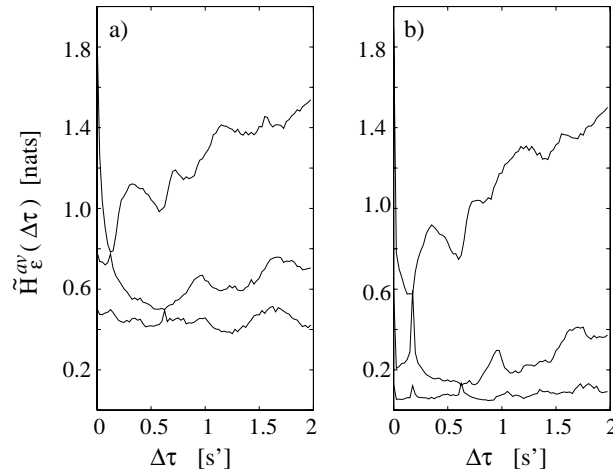


Abbildung 47: Die bedingten Entropien $H_\epsilon^{av}(\Delta\tau)$ ohne Korrektur nichtlinearer Einflüsse (links) und $\tilde{H}_\epsilon^{av}(\Delta\tau)$ (rechts). Die Entropien wurden bzgl. der x -Komponente des Lorenzsystems ohne Meßrauschen berechnet.

Im ersten Schritt wird das Verfahren auf die unverrauschte Zeitreihe angewendet. In Abbildung 47 ist zu erkennen, daß die Entropiefunktionen $H_\epsilon^{av}(\Delta\tau)$, in die keine Korrektur der durch nichtlineare Einflüsse verursachten Informationsverluste eingeht, und $\tilde{H}_\epsilon^{av}(\Delta\tau)$ qualitativ vergleichbar sind, also gleiche Extremstellen aufweisen. Die korrigierten bedingten Entropien sind jedoch erheblich geringer, sogar im Fall einer zweidimensionalen Einbettung, bei der die zukünftigen x -Werte nicht eindeutig bestimmt sind. Damit ist es insbesondere durch die unkorrigierte bedingte Entropie $H_\epsilon^{av}(Y|X)$ nicht möglich, die benötigte Einbettungsdimension direkt aus deren Verschwinden zu bestimmen.

Im nächsten Schritt wird normalverteiltes Rauschen auf die Zeitreihe addiert. Dies entspricht einem Meßrauschen bei experimentellen Daten. Die bedingten Entropiekurven werden dann bzgl. der neuen Zeitreihe berechnet. Das Rauschen wird mit einem Signal-Rausch-Verhältnis von 40 dB angesetzt, wobei das Signal-Rausch-Verhältnis durch den Quotienten aus den Standardabweichungen von Signal X und Rauschen ξ bestimmt ist

$$\Phi_{SNR}(X, \xi) = 20 \log_{10} \left(\frac{\Delta X}{\Delta \xi} \right).$$

Bei einer Wahl von $\epsilon = \frac{1}{50}$ entspricht die Standardabweichung bei $\Phi_{SNR}(X, \xi) = 40$ dB ungefähr $\Delta \xi \simeq \frac{\epsilon}{2}$.

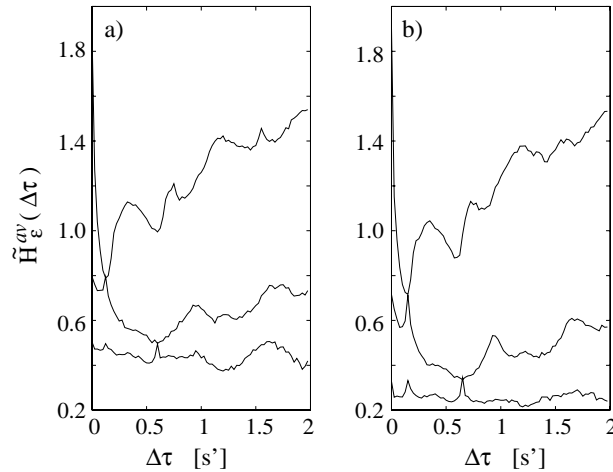


Abbildung 48: Die bedingten Entropien $H_\epsilon^{av}(\Delta\tau)$ ohne Korrektur nichtlinearer Einflüsse (links) und $\tilde{H}_\epsilon^{av}(\Delta\tau)$ (rechts) bzgl. der x -Komponente des Lorenzsystems, berechnet bei einem Signal-Rausch-Verhältnis von $\Phi_{SNR}(X, \xi) = 40$ dB und $\epsilon = \frac{1}{50}$.

Abbildung 48 zeigt die Kurvenverläufe der bedingten Entropien (analog zu Abbildung 47) für die verrauschte Zeitreihe. Auf die nichtkorrigierten Entropien haben die geringen Störungen keine wesentlichen Auswirkungen. Auch bei Berücksichtigung nichtlinearer Effekte (rechts) entspricht der Kurvenverlauf qualitativ dem in Abbildung 47 b) gezeigten. Allerdings sind die Entropiewerte erheblich höher, was nicht alleine mit dem durch das zusätzliche Rauschen verursachten Informationsverlust zu erklären ist. Vielmehr ist die Bestimmung der nichtlinearen Einflüsse nicht mehr ohne weiteres möglich, da in den lokalen ϵ -Umgebungen der Rauschanteil zu groß ist, um die linearisierte Abbildung exakt zu bestimmen. Übersteigt die Standardabweichung des Rauschens die Größe der ϵ -Umgebungen deutlich, haben die Korrekturfaktoren keinen Einfluß mehr auf die Werte von $\tilde{H}_\epsilon^{av}(\Delta\tau)$. Die Entropiekurven mit und ohne Korrektur sind dann identisch. Daher ist es sinnvoll, die ϵ -Umgebungen möglichst groß zu wählen. Dies hat, wie in Kapitel 4 gezeigt, auch den Vorteil, daß weniger Datenpunkte benötigt werden.

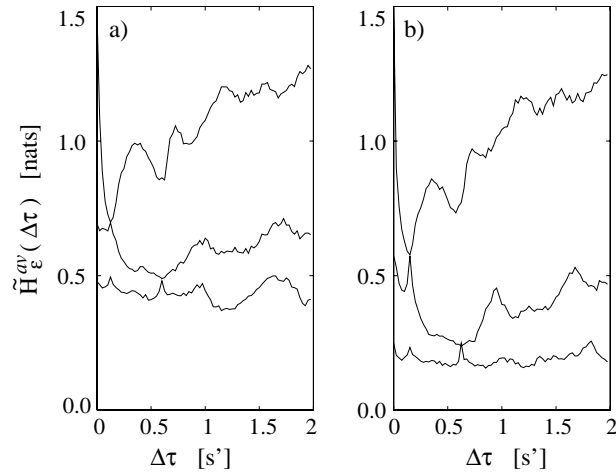


Abbildung 49: Die bedingten Entropien $H_\epsilon^{av}(\Delta\tau)$ ohne Korrektur nichtlinearer Einflüsse (links) und $\tilde{H}_\epsilon^{av}(\Delta\tau)$ (rechts) bzgl. der x -Komponente des Lorenzsystems, berechnet bei einem Signal-Rausch-Verhältnis von $\Phi_{SNR}(X, \xi) = 40$ dB und $\epsilon = \frac{1}{30}$.

Durch die Vergrößerung von ϵ tritt die Korrektur wieder deutlicher hervor (Abbildung 49). Die Abweichungen in den Absolutwerten sind dabei durch die ϵ -Abhängigkeit von $H_\epsilon^{av}(\Delta\tau)$ und $\tilde{H}_\epsilon^{av}(\Delta\tau)$ bedingt. Durch das größere ϵ lassen sich jedoch die linearisierten Abbildungen besser bestimmen, da der Rauscheinfluß deutlich verringert ist. Dadurch wird der relative Unterschied zwischen den

bedingten Entropien $\tilde{H}_\epsilon^{av}(\Delta\tau)$ und $H_\epsilon^{av}(\Delta\tau)$, auf welche die Wahl von ϵ keinen weiteren Einfluß hat, größer. Allerdings müssen die Umgebungen noch klein genug sein, um eine sinnvolle Abschätzung der linearisierten Abbildung zu gewährleisten. Wird das Rauschen zu groß, das Signal-Rausch-Verhältnis also zu klein, können die nichtlinearen Einflüsse nicht mehr bestimmt werden.

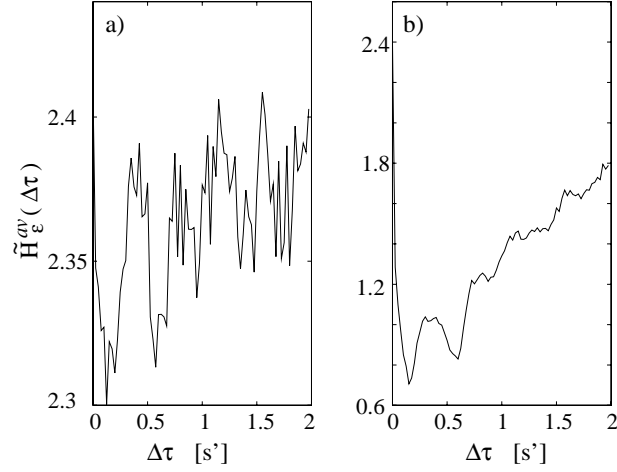


Abbildung 50: Die bedingte Entropie $\tilde{H}_\epsilon^{av}(\Delta\tau)$ bzgl. der x -Komponente des Lorenzsystems bei einem Signal-Rausch-Verhältnis von $\Phi_{SNR}(X, \xi) = 5$ dB (links) und ohne Rauschen (rechts) bei $\epsilon = \frac{1}{50}$.

Ein Signal-Rausch-Verhältnis von $\Phi_{SNR}(X, \xi) = 5$ dB bedeutet, daß die Standardabweichung des Rauschens ungefähr einem Drittel der Gesamtamplitude des Signals entspricht. Es ist also nicht mehr möglich, ϵ groß genug zu wählen. D.h. die nichtlineare Korrektur hat für die Berechnung der modifizierten bedingten Entropie keine Bedeutung mehr (Abbildung 50a). Die bedingte Entropie ist in diesem Fall aufgrund des durch das Rauschen verursachten Informationsverlusts sehr hoch. Dennoch weist $\tilde{H}_\epsilon^{av}(\Delta\tau)$ für die stark verrauschten Daten noch die gleichen Minimalstellen auf wie für die unverrauschte Zeitreihe. Die Berechnung der modifizierten bedingten Entropie ist also auch im Fall „schlechter“, stark verrauschter Daten noch sinnvoll: Qualitative Aussagen sind somit immer noch möglich. Die Quantifizierung der nichtlinearen Einflüsse wird jedoch mit abnehmendem Signal-Rausch-Verhältnis schwieriger.

10.2 Methoden zur Bestimmung von Abhängigkeiten innerhalb von Zeitreihen

In diesem Abschnitt werden andere Methoden zur Bestimmung von Zusammenhängen kurz beschrieben und mit dem in dieser Arbeit eingeführten Verfahren verglichen.

10.2.1 Bedingte und relative Entropie

Im letzten Abschnitt wurde gezeigt, daß die modifizierte bedingte Entropie $\tilde{H}_\epsilon(Y|X)$ bei Meßrauschen das gleiche Verhalten wie die nicht korrigierte Entropie

$$H_\epsilon(Y|X) = -E[\log P_\epsilon(y|x)] \quad (10.1)$$

aufweist. Allerdings werden durch (10.1) nichtlineare Informationsverluste nicht mehr von anderen unterschieden. So ergibt die bedingte Entropie für die logistische Parabel

$$H_\epsilon(Y|X) \simeq \log 2.$$

Dieses Ergebnis ist dadurch bedingt, daß an Stellen, an denen Steigungen $f'(x) > 1$ auftreten, die bedingte Wahrscheinlichkeit $P_\epsilon(y|x) < 1$ ist.

Diese Einflüsse können umgangen werden, wenn die „Wahrscheinlichkeit“ eines Punktes \mathbf{z}_n durch den mittleren Abstand $R^{(k)}(\mathbf{z}_n)$ der k nächsten Nachbarn von \mathbf{z}_n abgeschätzt wird. $R^{(k)}(\mathbf{z}_n)$ ist umgekehrt proportional zu $P_\epsilon(\mathbf{z}_n)$:

$$R^{(k)}(\mathbf{z}_n) = \frac{1}{k} \sum_{j=1}^k (\mathbf{z}_n - \mathbf{z}_{s_{n,j}})^2 \sim \frac{1}{P_\epsilon(\mathbf{z}_n)}, \quad (10.2)$$

wobei $s_{n,j}$ die Indices der k Nachbarnpunkte von \mathbf{z}_n sind.

Um Abhängigkeiten zwischen zwei Größen X und Y mit d_1 - bzw. d_2 -dimensionalen Zustandsvektoren zu bestimmen, wurde daher in [41] folgende Methode vorgeschlagen:

Zunächst wird gemäß (10.2) die „Wahrscheinlichkeit“, oder genauer die Dichte, des Punktes \mathbf{y}_n abgeschätzt. Danach werden zu dem korrespondierenden Punkt \mathbf{x}_n , zur gleichen Zeit n , ebenfalls die k nächsten Nachbarn $\mathbf{x}_{r_{n,j}}$ bestimmt. Damit kann eine Abschätzung für die relative Wahrscheinlichkeit durch

$$S_n^{(k)}(\mathbf{y}_n|\mathbf{x}_n) = \frac{R^{(k)}(\mathbf{y}_n)}{R^{(k)}(\mathbf{y}_n|\mathbf{x}_n)}$$

erfolgen, wobei $R_k(\mathbf{y}_n|\mathbf{x}_n)$ der mittlere Abstand der zu den k nächsten Nachbarn von \mathbf{x}_n korrespondierenden \mathbf{y} -Punkte $\mathbf{y}_{r_{n,j}}$ zu \mathbf{y}_n ist:

$$R^{(k)}(\mathbf{y}_n|\mathbf{x}_n) = \frac{1}{k} \sum_{j=1}^k (\mathbf{y}_n - \mathbf{y}_{r_{n,j}})^2. \quad (10.3)$$

Ist \mathbf{y}_n eineindeutig durch \mathbf{x}_n bestimmt, so ist

$$S_n^{(k)}(\mathbf{y}_n|\mathbf{x}_n) = 1,$$

andernfalls nimmt $S(\mathbf{y}_n|\mathbf{x}_n)$ einen Wert

$$0 \leq S_n^{(k)}(\mathbf{y}_n|\mathbf{x}_n) < 1$$

an. Somit können durch Mittelwertbildung Abhängigkeiten zwischen den Zufallsgrößen X und Y bestimmt werden. Mittels $S_n(\mathbf{y}_n|\mathbf{x}_n)$ kann auch die relative (Kullback-Leibler-) Entropie (Anhang B)

$$S^{(k)}(Y|X) = -E[\log S_n(\mathbf{y}_n|\mathbf{x}_n)] \quad (10.4)$$

definiert werden. Diese nimmt bei einer durch eine bijektive Abbildung gegebenen Abhängigkeit unabhängig von den Ableitungen der Funktion den Wert 0 an. Allerdings ist dies nicht mehr gegeben, wenn die \mathbf{y} -Werte verschiedene Urbilder besitzen können. In diesem Fall besitzen die nächsten Nachbarn von \mathbf{y}_n Urbilder, welche einen sehr großen Abstand von \mathbf{x}_n besitzen können und damit auf keinen Fall zu dessen Nachbarn gehören. Daraus folgt, daß $R_k(\mathbf{y}_n) < R_k(\mathbf{y}_n|\mathbf{x}_n)$ ist und $S(Y|X) > 0$, obwohl eine direkte Abhängigkeit vorliegt. Im Extremfall, wenn k die Anzahl der Urbilder von \mathbf{y} ist (im Fall der Parabel also $k = 2$), verschwindet die relative Wahrscheinlichkeit, sogar wenn Y eindeutig von X abhängt. Das bedeutet, die relative Entropie (10.4) wächst über alle Maßen. Ein Vorschlag zur Verbesserung des Verfahrens ist in diesem Fall $S_n^{(k)*}(\mathbf{y}_n|\mathbf{x}_n)$ durch das Maximum

$$S_n^{(k)*}(\mathbf{y}_n|\mathbf{x}_n) = \max(S_n(\mathbf{y}_n|\mathbf{x}_n), (R^{(k)}(\mathbf{x}_n))^{-1})$$

zu definieren.

10.2.2 Transinformation

Eine weit verbreitete Methode zur Bestimmung von Abhängigkeiten in Zeitreihen nichtlinearer Systeme ist die Berechnung der Transinformation [23, 27, 42]. Dazu können, wie in Kapitel 4 beschrieben wurde, die Wahrscheinlichkeiten $P_\epsilon(\mathbf{x})$,

$P_\epsilon(y)$ und $P_\epsilon(\mathbf{x}, y)$ berechnet werden. Die Transinformation ergibt sich dann gemäß (2.10) durch

$$I(X, Y) = E \left[\log \left(\frac{P_\epsilon(\mathbf{x}, y)}{P_\epsilon(\mathbf{x})P_\epsilon(y)} \right) \right]. \quad (10.5)$$

Analog zur Einführung verallgemeinerter Entropien (Anhang B) kann eine verallgemeinerte Transinformation

$$I_2(X, Y) = \log \left(\frac{\sum_{\alpha, \beta} P(Y = y_\alpha, X = x_\beta)^2}{\sum_\alpha P(Y = y_\alpha)^2 \sum_\beta P(X = x_\beta)^2} \right) \quad (10.6)$$

definiert werden. Diese ist jedoch nicht immer positiv und daher kein echtes Maß für Information [43]. Ist aber die Zufallsvariable Y gleichverteilt, so hat $I_2(X, Y)$ wieder die geforderte Eigenschaft [44]. Zur Berechnung der verallgemeinerten Transinformation werden daher zunächst die Datenpunkte y_t auf eine Gleichverteilung transformiert. Dies kann durch Sortieren der Daten nach ihrer Größe und anschließender Zuordnung der neuen Indices erfolgen. Diese Gleichverteilung ermöglicht eine schnelle Berechnung der Wahrscheinlichkeiten P_ϵ , da die nächsten Nachbarn eines Punktes durch das Sortieren schon bekannt sind und sich die Abstände aufgrund der Gleichverteilung direkt berechnen lassen.

Die Transinformationen (2.10) und (10.6) verschwinden, wenn X und Y voneinander statistisch unabhängig sind. Andernfalls nimmt sie positive Werte an; ist Y eine Funktion von X , so ist $I(X, Y) = H(Y)$. Damit lassen sich aber nur aufgrund positiver Transinformation keine Schlußfolgerungen über direkte Abhängigkeiten zwischen X und Y ziehen. Dazu müßte die Einzelentropie $H(Y)$ bekannt sein. Die Differenz

$$H(Y) - I(X, Y) = H(Y|X)$$

führt dann wieder zur bedingten Entropie.

Auch über die Berechnung mittlerer Abstände läßt sich ein der Transinformation entsprechendes Maß definieren [41]. Dazu werden die Abstände k_{rand} zufällig ausgewählter Punkte zu dem Referenzpunkt \mathbf{y}_n berechnet

$$R_n(\mathbf{y}_n) = \frac{1}{k} \sum_{j=1}^k (\mathbf{y}_n - \mathbf{y}_{k_{rand}})^2.$$

Die Größe

$$\tilde{I}^{(k)} = E \left[\log \frac{R_n(\mathbf{y}_n)}{R^{(k)}(\mathbf{y}_n|\mathbf{x}_n)} \right],$$

wobei der Nenner durch (10.3) gegeben ist, verschwindet, wenn X und Y unabhängig voneinander sind, und ist positiv, wenn eine Abhängigkeit besteht. Allerdings kann sie auch negative Werte annehmen. Da sie außerdem nicht symmetrisch in X und Y ist, stellt sie im eigentlichen Sinne keine Transinformation dar, kann aber dennoch in der Regel äquivalent verwendet werden.

10.2.3 Kovarianz

Lineare Methoden zur Identifizierung von Zusammenhängen sind in der Zeitreihenanalyse weit verbreitet. Da sie nicht auf informationstheoretischen Ansätzen beruhen und daher keine Abschätzung von Wahrscheinlichkeiten oder Dichten erforderlich ist, werden für diese Methoden relativ wenig Daten benötigt. Im wesentlichen basieren die Methoden auf der Berechnung der Kovarianz

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]. \quad (10.7)$$

Sind die Zufallsvariablen X und Y voneinander unabhängig, so ist $Cov(X, Y) = 0$. Ist $Cov(X, Y) > 0$, so bedeutet das, daß Y bei wachsendem X im Mittel größer wird, bei negativer Kovarianz wird Y in diesem Fall kleiner. Um ein festes Maß für lineare Abhängigkeiten zu erhalten, wird die Kovarianz durch das Produkt der Standardabweichungen $\Delta X \Delta Y$ dividiert. Die auf diese Weise erhaltene Größe ist die Korrelation

$$Corr(X, Y) = \frac{Cov(X, Y)}{\Delta X \Delta Y}. \quad (10.8)$$

Geht Y durch eine lineare Abbildung aus X hervor, so ist $|Corr(X, Y)| = 1$. Sind X und Y voneinander unabhängig, so ist $Corr(X, Y) = 0$. Allerdings folgt aus dem Verschwinden der Korrelation nicht die statistische Unabhängigkeit von X und Y , $P(Y|X) = P(Y)$. So ergibt z.B. die Korrelation bei der Untersuchung eines Kreises den Wert 0, obwohl jedem x -Wert nur maximal zwei y -Werte entsprechen, also $P(Y|X) = \frac{1}{2} \gg P(Y)$ ist. Dennoch ist die Korrelation aufgrund ihrer schnellen Berechenbarkeit eine der wichtigsten Größen zur Identifizierung von Abhängigkeiten. Insbesondere zeitliche Zusammenhänge, welche durch Kreuz- bzw. Autokorrelationen angegeben werden, lassen sich unter Verwendung des Korrelations- bzw. Wiener-Khintchine-Satzes sehr schnell bestimmen. Der Korrelationssatz besagt, daß das Produkt der Fouriertransformierten einer Größe $x(t)$ mit dem konjugiert Komplexen der Fouriertransformierten einer weiteren Größe

$y(t)$ der Fouriertransformierten ihrer Kreuzkorrelation $Corr(X_{\Delta t}, Y_0)$ entspricht. Der Wiener-Khintchine-Satz ist die Übertragung des Korrelationssatzes auf den Fall $X = Y$. Die Autokorrelationsfunktion ist damit durch die Fouriertransformierte des Leistungsspektrums gegeben. Mittels der schnellen Fouriertransformation (FFT) lassen sich die Korrelationsfunktionen in sehr kurzer Zeit berechnen.

Räumliche Abhängigkeiten lassen sich durch die Hauptkomponentenanalyse (oder auch Karhunen-Loève-Verfahren) bestimmen. Dazu werden zu jedem Paar von Meßreihen X_i und X_j an verschiedenen Orten i, j die Kovarianzen $Cov(X_i, X_j)$ bestimmt. Eine Singulärwertzerlegung der so bestimmten Kovarianzmatrix $C = (Cov(X_i, X_j))_{i,j}$ in

$$C = U W V$$

ermöglicht, die Daten auf ihre Hauptachsen zu transformieren. Die quadratische Matrix V bewirkt dabei die lineare Transformation auf die neuen Achsen, die Diagonalmatrix W beinhaltet die Singulärwerte. Die Raumrichtung mit dem größten Singulärwert zeigt dabei in Richtung der größten Kovarianz. Richtungen mit sehr kleinem Singulärwert sind für die Beschreibung des Systems unerheblich. Die Vernachlässigung dieser Raumrichtungen führt zu einer Dimensionsreduktion. Das System ist weitgehend durch die verbleibenden Hauptachsen bestimmt.

10.2.4 Zusammenfassung

Keines der hier vorgestellten Verfahren erlaubt es, nichtlineare Abhängigkeiten in Zeitreihen eindeutig zu identifizieren. Allerdings ist dies, wie im vorangegangenen Abschnitt gezeigt, bei Daten mit Meßrauschen auch durch die Berechnung von $\tilde{H}_\epsilon(Y|X)$ nicht ohne weiteres möglich. Daher hängt es weitgehend von der Beschaffenheit der vorhandenen Daten, sowie dem Zweck der Analyse ab, welche Methode am besten geeignet ist.

Bei kurzen, verrauschten Zeitreihen ist häufig nur möglich, lineare Methoden zu verwenden. Diese haben außerdem den Vorteil, daß die entsprechenden Größen sehr schnell zu berechnen sind. Nichtlineare Zusammenhänge werden jedoch nicht berücksichtigt. Insbesondere können somit auch keine einseitigen Abhängigkeiten identifiziert werden, da Korrelationen symmetrische Maße sind.

Ebenfalls symmetrisch ist die Transinformation, welche aber auch Nichtlinearitäten berücksichtigt. Durch ihr Verschwinden bei Unabhängigkeit der Meß-

größen ist sie besonders geeignet, Größen mit möglichst wenig gemeinsamer Information zu ermitteln.

Bedingte und relative Entropien sind asymmetrische Maße. Daher eignen sie sich gut, um Abhängigkeiten zwischen Zufallsgrößen zu beschreiben. Die bedingte Entropie $H_\epsilon(Y|X)$ nimmt jedoch bei nichtlinearen Abhängigkeiten und die relative Entropie $S^{(k)}(Y|X)$ bei nicht eindeutigen Abhängigkeiten zu große Werte an. Eine Kombination beider Größen, bei welcher eine bedingte Wahrscheinlichkeit durch

$$P^{(k)} = \frac{R^{(k)}(\mathbf{x}_n)}{R^{(k)}(\mathbf{x}_n | (\mathbf{x}_n, \mathbf{y}_n))},$$

definiert wird, wobei $R^{(k)}(\mathbf{x}_n | (\mathbf{x}_n, \mathbf{y}_n))$ der mittlere Abstand der zu den k nächsten Nachbarn von $(\mathbf{x}_n, \mathbf{y}_n)$ korrespondierenden \mathbf{x} -Werte ist, kann für verrauschte Zeitreihen geeignet sein. Allerdings können dabei die nichtlinearen Einflüsse nicht mehr getrennt von den anderen Informationsverlusten betrachtet werden.

Zur Beschreibung der Abhängigkeiten in nichtlinearen Systemen mit geringem Rauschanteil ist die in dieser Arbeit vorgestellte modifizierte bedingte Entropie $\tilde{H}_\epsilon(Y|X)$ am besten geeignet.

11 Zusammenfassung

In dieser Arbeit wurde gezeigt, welchen Beitrag die Berechnung bedingter Entropien aus Zeitreihen zur Identifizierung und damit zum Verständnis von Abhängigkeiten zwischen verschiedenen Meßgrößen und Informationsflüssen innerhalb einer Meßreihe leisten kann. Sie eignet sich zur Vorhersage physikalischer Größen und kann bei deren Modellierung verwendet werden.

Dazu wurde eine modifizierte bedingte Entropie eingeführt, welche auch durch Nichtlinearitäten verursachte Informationsverluste berücksichtigt. Es wurde gezeigt, wie der Informationsverlust bzgl. einer einzelnen Beobachtungsgröße in einem höherdimensionalen System quantifiziert werden kann. Dabei wurde insbesondere der Fall, daß ein Systemzustand nur durch die Messung festgelegter Größen bestimmt werden kann, untersucht.

Es wurde daraufhin ein Verfahren vorgestellt, mit dem diese modifizierte bedingte Entropie aus Zeitreihen berechnet werden kann. Dadurch wird ermöglicht, Abhängigkeiten zu quantifizieren, optimale Größen zur Beschreibung eines Systems zu ermitteln und festzustellen, wann eine Meßgröße vollständig durch andere beschrieben werden kann. Dieses Verfahren wurde zur Untersuchung unterschiedlicher Fragestellungen eingesetzt.

Durch die Asymmetrie der bedingten Entropie und die Berücksichtigung nichtlinearer Effekte eignet sich die beschriebene Methode sehr gut zur Bestimmung zeitlicher Abhängigkeiten in Zeitreihen nichtlinearer dynamischer Systeme, wie am Beispiel des Lorenzsystems gezeigt werden konnte. Die Eigenschaft der modifizierten bedingten Entropie, unter Hinzunahme schon bekannter Information nicht geringer zu werden, ermöglicht die Wahl der bestmöglichen Verzögerungszeiten für eine Attraktorrekonstruktion. Darüber hinaus konnte gezeigt werden, daß es günstig sein kann, verschiedene Verzögerungszeiten zu verwenden, wenn das dynamische System mehrere charakteristische „Frequenzen“ aufweist. Auf diese Weise kann die benötigte Einbettungsdimension reduziert werden.

Weiterhin wurde die modifizierte bedingte Entropie zur Untersuchung raumzeitlicher Zusammenhänge verwendet. Es konnte gezeigt werden, daß in Ketten gekoppelter nichtlinearer Abbildungen Informationsflüsse stattfinden, die nicht direkt den entsprechenden Entwicklungsgleichungen entnommen werden können.

Dennoch ist es in vielen Fällen möglich, die für die Dynamik des Systems wesentlichen Abhängigkeiten zu identifizieren.

Es gelang, in chaotisch getriebenen Systemen nichtlineares Resonanzverhalten nachzuweisen. Dieses hängt einerseits von der Kopplungsamplitude, andererseits und besonders von dem Verhältnis der Phasenraumgeschwindigkeiten des treibenden und getriebenen Systems ab.

Bei experimentellen Daten wurde der Vorteil des beschriebenen Verfahrens gegenüber linearen Standardmethoden demonstriert. So konnten bei Zeitreihen eines Halbleiterlasers mit zwei externen Resonatoren die Umlaufzeiten in diesen Resonatoren klarer bestimmt werden. Dadurch wurde ermöglicht, eine Verstimmung zwischen der Grundfrequenz und ihrer dritten Harmonischen zu identifizieren. Durch Messungen an einem quasiperiodisch getriebenen, nichtlinearen Schwingkreis und die Berechnung der modifizierten bedingten Entropie aus den resultierenden Zeitreihen wurde die Aussagekraft dieser Ergebnisse bestätigt.

Schließlich wurde eine Vorgehensweise vorgeschlagen, mit der aus multivariaten Zeitreihen ein optimales datengetriebenes Modell erstellt werden kann. Die Vorteile, die dabei eine Berechnung der bedingten Entropie bietet, wurden am Beispiel der Modellbildung an Turbogeneratoren demonstriert. Das Verfahren wurde in ein im Rahmen dieser Arbeit erstelltes Programmpaket zur nichtlinearen Zeitreihenanalyse in Echtzeit integriert. Dieses erlaubt die parallele Analyse mit verschiedenen Verfahren, sowie die Kombination dieser Analysemethoden.

Es wurde anhand von zahlreichen Beispielen gezeigt, daß die modifizierte bedingte Entropie hervorragend zur Bestimmung von Abhängigkeiten in und zwischen Zeitreihen geeignet ist. Jedoch können durch ihre Berechnung nicht direkte Ursache-Wirkung-Relationen identifiziert werden; sie kann nur Hinweise auf Zusammenhänge geben, die anschließend physikalisch interpretiert werden müssen.

Bei zahlreichen bisher unverstandenen Phänomenen kann die Berechnung der bedingten Entropie hilfreiche Anregungen geben. In vielen komplexen Systemen können zeitliche Zusammenhänge nicht mehr durch lineare Methoden ermittelt werden. Durch die in dieser Arbeit beschriebenen Vorteile kann die bedingte Entropie nunmehr in diesem Rahmen gut eingesetzt werden.

Bei räumlich ausgedehnten dynamischen Systemen kann ihre Berechnung sehr sinnvoll sein. So ist z.B. ihr Einsatz bei der Beschreibung der Dynamik in Breit-

streifenlasern oder bei der Untersuchung von EEG-Daten denkbar. Insbesondere für die Untersuchung verallgemeinerter Synchronisationsphänomene scheint die modifizierte bedingte Entropie hervorragend geeignet zu sein.

Eine in Kapitel 10 angedeutete Modifikation könnte die Entropieberechnung für experimentelle Daten mit starkem Rauscheinfluß möglicherweise noch effizienter gestalten.

A Wahrscheinlichkeitstheoretische Grundlagen

In diesem Anhang werden einige Grundbegriffe der Wahrscheinlichkeitstheorie, welche für die Zeitreihenanalyse von fundamentaler Bedeutung sind, kurz eingeführt [45, 9, 46, 1].

Wahrscheinlichkeitsmaße und Dichten

Definition A.1 (σ -Algebra) Gegeben sei eine beliebige Menge Ω . Eine σ -Algebra \mathcal{A} ist ein System von Teilmengen von Ω , das folgende Eigenschaften besitzt:

1. $\Omega \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow \complement A \in \mathcal{A}$,
3. für jede Folge $(A_n)_{n \in \mathbb{N}}$ von Mengen aus \mathcal{A} liegt $\bigcup_{n=1}^{\infty} A_n$ in \mathcal{A} . Dabei ist $\complement A$ das Komplement von A in Ω .

Definition A.2 (Maß) Sei (Ω, \mathcal{A}) ein Meßraum und $A_n \in \mathcal{A}$. Eine auf \mathcal{A} definierte Funktion μ mit Werten in \mathbb{R}_0^+ heißt ein Maß (auf \mathcal{A}), wenn sie folgende Eigenschaften besitzt:

$$\mu(\emptyset) = 0, \tag{A.1}$$

und für jede Folge (A_n) paarweise disjunkter Mengen aus \mathcal{A} gilt:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{A.2}$$

Das translationsinvariante Maß, das jedem Einheitwürfel aus dem \mathbb{R}^d den Wert 1 zuordnet, heißt Lebesguemaß.

Definition A.3 (Wahrscheinlichkeitsmaß) Ein Maß P auf \mathcal{A} heißt Wahrscheinlichkeitsmaß, wenn gilt:

$$P(\Omega) = 1. \tag{A.3}$$

Definition A.4 (Maßraum) Das Tripel $(\Omega, \mathcal{A}, \mu)$ heißt Maßraum. Ist $\mu = P$ ein Wahrscheinlichkeitsmaß, so heißt (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Satz A.5 $(\Omega, \mathcal{A}, \mu)$ sei ein Maßraum und $\rho \geq 0$ eine beliebige meßbare¹⁰, numerische Funktion auf Ω . Dann wird durch

$$\nu(A) := \int_A \rho \, d\mu \quad (\text{A.4})$$

wiederum ein Maß auf \mathcal{A} definiert.

(Beweis in [45])

Definition A.6 (Maße mit Dichten) Ist ρ eine nichtnegative, meßbare Funktion auf Ω , so heißt das Maß ν aus (A.4) das Maß mit der Dichte ρ (bzgl. μ).

Zufallsvariablen und stochastische Prozesse

Definition A.7 (Zufallsvariable) Sei $(\Omega_1, \mathcal{A}_1, P)$ ein Wahrscheinlichkeitsraum und $(\Omega_2, \mathcal{A}_2)$ ein Meßraum. Jede \mathcal{A}_1 - \mathcal{A}_2 -meßbare Abbildung $X : \Omega_1 \rightarrow \Omega_2$ heißt Zufallsvariable.

In dieser Arbeit werden hauptsächlich reellwertige Zufallsvariablen betrachtet, d.h. $\Omega_2 = \mathbb{R}$. Die Abbildung $X : \Omega_1 \rightarrow \mathbb{R}^d$ ist genau dann eine Zufallsvariable, wenn jede Komponente eine reellwertige Zufallsvariable ist.

Definition A.8 (Stochastischer Prozeß) Gegeben sei eine Indexmenge T , welche die Zeit repräsentiert. Ein stochastischer Prozeß ist eine Familie von Zufallsvariablen $\{X_t\}_{t \in T}$ auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Er wird daher durch das Quadrupel $(\Omega, \mathcal{A}, P, X_{t \in T})$ dargestellt.

¹⁰Eine Funktion $T : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ heißt meßbar, wenn die Urbilder aller Mengen $A_i \in \mathcal{A}_2$ in der σ -Algebra \mathcal{A}_1 enthalten sind.

B Wichtige Begriffe aus Informations- und Ergodentheorie

B.1 Ergodentheoretische Begriffe

Ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) sei gegeben. Ein durch $\{f^t\}$ gegebenes dynamisches System heißt *ergodisch*, wenn für jede Menge $A \in \mathcal{A}$ mit $f^{-t}(A) = A$ entweder $P(A) = 0$ oder $P(A) = 1$ gilt.

In ergodischen Systemen kann der Erwartungswert einer Funktion h durch das zeitliche Mittel ersetzt werden

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int h(f^t(x)) dt = \int h(x) dP(x).$$

In zeitdiskreten dynamischen Systemen heißt ein invariantes Maß μ *Sinai-Ruelle-Bowen-Maß* (oder kurz SRB-Maß), $\mu = \mu_{SRB}$, wenn für jede stetige Funktion $h : \Omega \rightarrow \mathbb{R}$ die Menge aller der Punkte $x_0 \in \Omega$, für die

$$\int_{\Omega} h d\mu_{SRB} = \lim_{N_T \rightarrow \infty} \frac{1}{N_T} \sum_{t=1}^{N_T} h(f^t(x_0)),$$

gilt, ein positives Lebesguemaß hat [47].

B.2 Informationstheoretische Begriffe

Neben der Shannon-Entropie werden in der Informationstheorie auch andere Maße verwendet.

Die *Kullback-Leibler-Entropie* oder *Divergenz* [48] ist durch den Quotienten zweier Verteilungen bestimmt

$$K(P^{(1)}|P^{(0)}) = \sum_{\alpha=1}^N P_{\alpha}^{(1)} \log \left(\frac{P_{\alpha}^{(1)}}{P_{\alpha}^{(0)}} \right). \quad (\text{B.1})$$

Sie ist positiv und nicht symmetrisch. $K(P^{(1)}|P^{(0)})$ läßt sich als Abstandsmaß zwischen zwei Verteilungen interpretieren und wird daher auch bisweilen *Kullback-Leibler-Abstand* genannt. Sie verschwindet, wenn $P^{(1)} = P^{(0)}$. Die Größe $S^{(k)}(Y|X)$ aus Gl. (10.4) kann als Kullback-Leibler-Entropie angesehen werden.

Eine Verallgemeinerung der Shannon-Entropie stellen die *Rényi-Entropien* [43] q -ter Ordnung dar

$$H_q(X) = \frac{1}{1-q} \log \left(\sum_{i=\alpha}^N P^q(X = x_\alpha) \right). \quad (\text{B.2})$$

Für $q = 1$ ergibt sich die Shannon-Entropie. Die Rényi-Entropien werden zur Definition verallgemeinerter Dimensionen

$$D_q = \lim_{\epsilon \rightarrow 0} \frac{H_{\epsilon,q}(X)}{-\log \epsilon}$$

verwendet. Die in dieser Arbeit verwendeten Informations- und Kapazitätsdimensionen ergeben sich als D_1 und D_0 . Ebenso können verallgemeinerte Transinformationen wie in (10.6) definiert werden.

C Echtzeitanalyse – der NLyzer

Im Rahmen dieser Arbeit wurde ein komplexes Werkzeug zur Zeitreihenanalyse in Echtzeit – der *NLyzer* – erstellt [37].

In den NLyzer sind zahlreiche Methoden integriert. Diese reichen von statistischen Verfahren über die linearen Standardmethoden, wie FFT und Korrelation bis zu modernen nichtlinearen Analysemethoden, wie sie in dieser Arbeit vorgestellt wurden.

C.1 Aufbau des NLyzers

Der NLyzer ist ein prozeßorientiertes Programmpaket. Alle enthaltenen Verfahren sind als eigenständige Programme implementiert. Diese wurden in C geschrieben; allerdings ist auch eine Implementierung in anderen Hochsprachen möglich, da jedes Programm eine vom NLyzer unabhängig ausführbare Einheit darstellt.

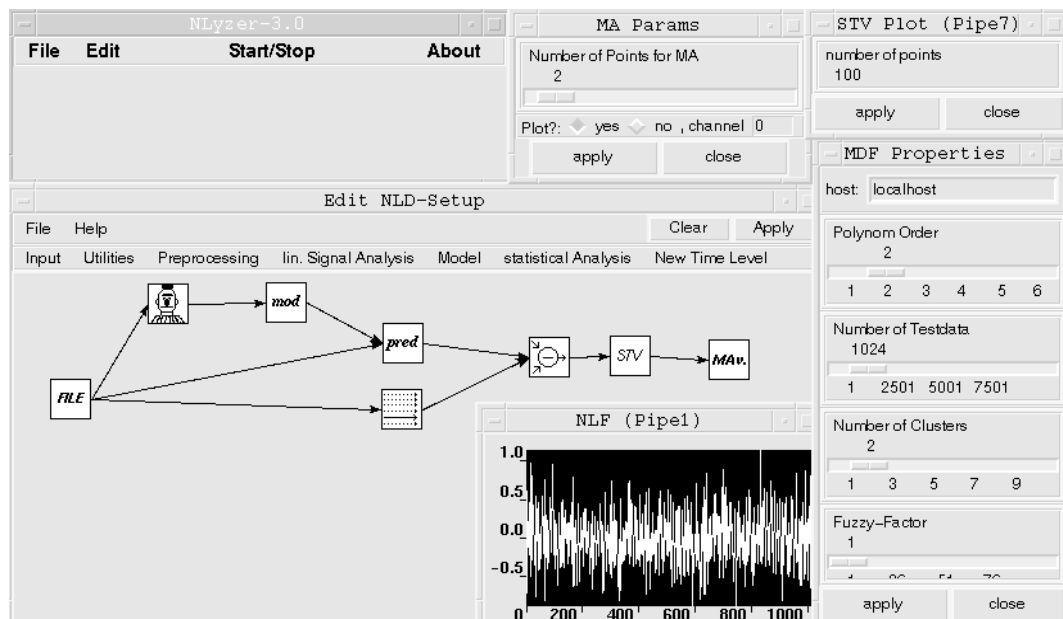


Abbildung 51: Verschaltung der Methoden im graphischen Editor.

Die enthaltenen Verfahren können mittels eines graphischen Editors miteinander verschaltet werden (Abbildung 51). Das in TCL/TK und C implementierte Hauptprogramm beinhaltet einerseits die Benutzer-Schnittstelle, andererseits die Steuerung des Analyseprozesses.

Die einzelnen Verfahren werden als Prozesse in einer Kette von Pipes gestartet. Dadurch ist eine Parallelisierbarkeit der Analysen gewährleistet. Die Steuerung erfolgt eingabeorientiert: mittels eines File-Event-Handlers wird erkannt, wann eine Pipe les- oder beschreibbar ist. Die für den Programmablauf und die Analyse notwendigen Parameter, sowie die Daten werden auf den Standardeingang der betreffenden Pipe geschrieben, die Ergebnisse vom Standardausgang gelesen. Durch diesen Aufbau ist es möglich, verschiedene Zeitebenen gleichzeitig zu untersuchen. So können auf kleinen Zeitskalen Kenngrößen, wie Momente oder fraktale Dimensionen berechnet werden, und auf einer höheren Zeitebene deren zeitlicher Verlauf analysiert werden. Weiterhin können sehr rechenaufwendige, zeitintensive Verfahren vom restlichen Analysevorgang losgelöst werden, um diesen nicht aufzuhalten. Die Dateneingabe kann wahlweise in einzelnen Datensätzen oder kontinuierlich erfolgen.

Zu jedem Analyseprozeß können die notwendigen Parameter individuell gewählt werden, d.h. identische Programme können gleichzeitig mit unterschiedlichen Parametern verwendet werden. Dazu werden die Analyse-Parameter in den jeweiligen graphischen Fenstern eingestellt (Abbildung 51). In diesen kann auch gewählt werden, ob die einzelnen Analyseergebnisse graphisch dargestellt werden sollen; dabei sind verschiedene Darstellungsformen möglich.

C.2 Beispiel: Modellbildung bei Motorlüftern

Am Beispiel der Modellbildung bei Motorlüftern soll der Aufbau eines Analyseprozesses demonstriert werden.

Ziel der Analyse ist es, defekte Lüfter von intakten zu unterscheiden. Dazu werden Körperschallsignale verwendet. Die defekten Lüfter weisen ein periodisches „Schlagen“ auf (Zeitreihe in Abb. 52, unten links), wobei die entsprechenden Spitzen bei verschiedenen Lüftern mit unterschiedlicher Amplitude auftreten. Daher werden die tiefpaßgefilterten Daten auf eine Gleichverteilung transformiert (Zeitreihe in Abb. 52, unten Mitte) und die so erhaltenen Daten eingebettet. Aus den nun mehrdimensionalen Zeitreihen intakter Lüfter wird ein Modell gebildet, daß zukünftige Zustände vorhersagt. Ist ein derartiges Modell gefunden, kann die Modellbildung abgeschaltet werden.

Im nächsten Schritt werden die Modellfehler berechnet. Dies geschieht mit

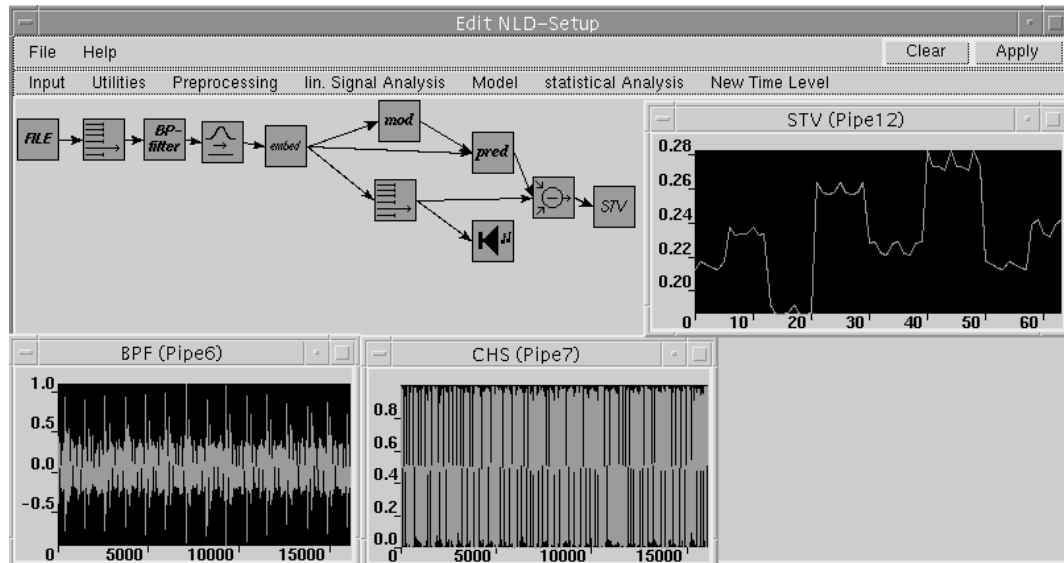
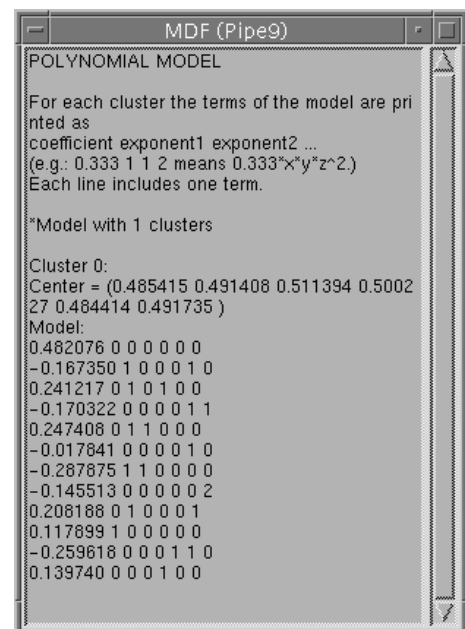


Abbildung 52: Aufbau zur Analyse und Modellbildung bei Motorlüftern.

Datensätzen der Länge 20000 (Datenpunkte). Der gemittelte Fehler kann zur Klassifizierung der Lüfter in defekt und intakt verwendet werden. Defekte Lüfter weisen einen deutlich höheren Modellfehler auf (Zeitreihe in Abb. 52, oben rechts). Zur Kontrolle können die Signale auch akustisch wiedergegeben werden.

Neben den Modellfehlern kann während der Modellbildung auch das Modell selbst überwacht werden. Die Anzeige des aktuellen Modells geschieht mittels eines eigenen Textfensters. In diesem Fenster werden die Art des Modells, Anzahl und Zentren der verwendeten Cluster und das Modell selbst angezeigt. Hierbei werden die zugrundeliegenden Basisfunktionen mit den zugehörigen Koeffizienten angegeben (Abb. rechts).



D Symbolverzeichnis

In dieser Arbeit wurden, wenn nicht ausdrücklich anders erwähnt, die folgenden Symbole verwendet.

Mathematische Symbole

Symbol	Bedeutung
Ω	beliebige Menge
M	Teilraum des Zustandsraums
\mathcal{A}	σ -Algebra
μ	beliebiges Maß
P	Wahrscheinlichkeitsmaß
ρ	Wahrscheinlichkeitsdichte
μ_L^d	d -dimensionales Lebesguemaß
X, Y	Zufallsvariablen
x_α, y_β	diskrete Zustände
$x, y, \mathbf{x}, \mathbf{y}$	stetige Zustände (skalar bzw. vektoriell)
$E(X)$	Erwartungswert von X
$H(X)$	(Shannon-)Entropie von X
$H(Y X)$	bedingte Entropie von Y bei gegebenen X
$h(Y X)$	bedingte Shannon-Entropie (stetiger Fall)
$h(P)$	Kolmogorov-Sinai-Entropie
$h^*(f)$	Informationsverlust in einer Meßrichtung
$\tilde{h}(Y X)$	modifizierte bedingte Entropie
$I(X, Y)$	Transinformation
$(\Omega, \mathcal{A}, P, X_{t \in T})$	stochastischer Prozeß
$\{x_t\}_{t \in T}$	Zeitreihe
$\mathfrak{M}, \mathfrak{N}$	Partitionen
\mathfrak{M}_ϵ	Partition in Mengen der Größe ϵ
$U_\epsilon(\mathbf{x})$	ϵ -Umgebung von \mathbf{x}
$B_\epsilon(\mathbf{x})$	Kugel mit Radius ϵ und Mittelpunkt \mathbf{x}
$C_\epsilon(\mathbf{x})$	Würfel mit Kantenlänge 2ϵ und Mittelpunkt \mathbf{x}

Literaturverzeichnis

- [1] P. BROCKWELL, R. DAVIS. *Time Series: Theory and Methods*. Springer-Verlag, New York, Berlin, Heidelberg (1991).
- [2] H. SCHUSTER. *Deterministic Chaos*. VCH Verlagsgesellschaft, Weinheim (1989).
- [3] J. ECKMANN, D. RUELLE. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, **57**(3), 617 – 656 (1985).
- [4] V. OSELEDEC. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, **19**, 197 – 231 (1968).
- [5] J. HOLZFUSS, U. PARLITZ. Lyapunov exponents from time series. In L. Arnold, H. Crauel, J.-P. Eckman, Herausgeber, *Lyapunov Exponents*, Band 1486 der *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, New York, Heidelberg, S. 263 – 270 (1991).
- [6] C. SHANNON, W. WEAVER. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana (1964).
- [7] R. GRAY. *Entropy and Information Theory*. Springer-Verlag, New York, Berlin, Heidelberg (1990).
- [8] P. WALTERS. *An Introduction to Ergodic Theory*. Springer-Verlag, New York, Berlin, Heidelberg (1982).
- [9] H. BAUER. *Wahrscheinlichkeitstheorie*. Walter de Gruyter, Berlin, New York (1991).
- [10] Y. PESIN. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Math. Survey*, **32**(4), 55–114 (1977).
- [11] M. BARNER, F. FLOHR. *Analysis II*. Walter de Gruyter, Berlin, New York (1989).

- [12] O. GÖHRS, J. HOLZFUSS. Conditional Entropy of Nonlinear Dynamical Systems. In D. S. Broomhead, E. A. Luchinskaya, P. V. E. McClintock, T. Mullin, Herausgeber, *Stochaos: Stochastic and Chaotic Dynamics in the Lakes*. American Institute of Physics, Woodbury, New York. in Druck (2000).
- [13] W. H. PRESS, S. A. TEUKOLKY, W. T. VETTERLING, B. P. FLANNERY. *Numerical Recipes in C*. Cambridge University Press, second edition (1992).
- [14] P. GRASSBERGER, I. PROCACCIA. Dimensions and entropies of strange attractors from a fluctuating dynamics approach. *Physica D*, **13**, 34–54 (1984).
- [15] J. D. FARMER, E. OTT, J. A. YORKE. The dimension of chaotic attractors. *Physica D*, **7**, 153–180 (1983).
- [16] G. ORTEGA. Invariant Measures as Lagrangian Variables: Their Application to Time Series Analysis. *Phys. Rev. Lett.*, **77**(2), 259 – 262 (1996).
- [17] G. ORTEGA. Smoothness Implies Determinism in Time Series: A Measure Based Approach. *Phys. Rev. Lett.*, **81**(20), 4345 – 4348 (1998).
- [18] F. TAKENS. Detecting strange attractors in turbulence. In D. Rand, L. Young, Herausgeber, *Dynamical Systems and Turbulence*, Band 898 der *Lecture Notes in Mathematics*. Springer-Verlag, S. 366 – 381 (1981).
- [19] T. SAUER, J. YORKE, M. CASDAGLI. Embedology. *J. Stat. Phys.*, **65**(3,4), 579 – 616 (1991).
- [20] P. COLLET, J.-P. ECKMANN. *Iterated Maps on the Interval as Dynamical System*. Birkhäuser, Basel (1980).
- [21] B. HUNT, T. SAUER, J. YORKE. Prevalence: A Translation-Invariant “Almost Every“ On Infinite Dimensional Spaces. *Bull. Am. Math. Soc.*, **27**(2), 217 – 238 (1992).
- [22] W. LIEBERT, K. PAWELZIK, H. G. SCHUSTER. Optimal embeddings of chaotic attractors from topological considerations. *Europhys. Lett.*, **14**(6), 521 (1991).

- [23] A. M. FRASER, H. L. SWINNEY. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, **33**(2), 1134–1140 (1986).
- [24] W. LIEBERT, H. G. SCHUSTER. Proper choice of the time delay for the analysis of chaotic time series. *Phys. Lett. A*, **142**(2-3), 107–111 (1989).
- [25] T. BUZUG, G. PFISTER. Comparison of algorithms calculating optimal parameters for delay time coordinates. *Physica D*, **58**, 127 (1992).
- [26] T. BUZUG, G. PFISTER. Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global and local dynamical behavior of strange attractors. *Phys. Rev. A*, **45**, 7073–7084 (1992).
- [27] J. HOLZFUSS, W. LAUTERBORN. Nonlinear Dynamics of Acoustic Cavitation Noise. In M. H. und D.T. Blackstock, Herausgeber, *Frontiers of Nonlinear Acoustics: Proceedings of the ISNA*. Elsevier Science, S. 464 – 469 (1990).
- [28] E. OTT, T. SAUER, J. YORKE, Herausgeber. *Coping With Chaos*, chapter 5. The Theory of Embedding. John Wiley & Sons, Inc., S. 44 – 50 (1994).
- [29] H. D. I. ABARBANEL, N. F. RULKOV, M. M. SUSHCHIK. Generalized synchronization of chaos: The auxiliary system approach. *Phys. Rev. E*, **53**(5), 4528–4535 (1996).
- [30] B. R. HUNT, E. OTT, J. A. YORKE. Differentiable generalized synchronization of chaos. *Phys. Rev. E*, **55**(4), 4029–4034 (1997).
- [31] K. IKEDA, K. MATSUMOTO. High-Dimensional Chaotic Behaviour in Systems with Time-Delayed Feedback. *Physica D*, **29**, 223 – 235 (1987).
- [32] I. FISCHER, O. HESS, W. ELSÄSSER, E. GÖBEL. High-Dimensional Chaotic Dynamics of an External Cavity Semiconductor Laser. *Phys. Rev. Lett.*, **73**(16), 2188 – 2191 (1994).
- [33] M. ORR. Introduction to Radial Basis Function Networks. Technical report, Centre for Cognitive Science, University of Edinburgh (1996).

- [34] R. ZÖLLER, F. REINE, V. TEWS. Zustandsanalyse rotierender Maschinen mittels Zeitreihenanalyse. In *Zeitreihenanalyse*. VDI-Technologiezentrum Physikalische Technologien, Düsseldorf, S. 9 – 12 (1997).
- [35] M. WINDHAM. Geometrical fuzzy clustering algorithms. *Fuzzy Sets and Systems*, **10**, 271 – 279 (1983).
- [36] R. ZÖLLER. *Methoden der Nichtlinearen Dynamik zur Zustandsanalyse technischer Systeme*. Dissertation, TH Darmstadt (1997).
- [37] O. GÖHRS, J. HOLZFUSS. NLyzer – Nichtlineare Analyse in Echtzeit. In *Technische Anwendungen von Erkenntnissen der Nichtlinearen Dynamik*. VDI-Technologiezentrum Physikalische Technologien, Düsseldorf, S. 327 – 330 (1999).
- [38] T. SCHREIBER. Efficient Neighbor Searching in Nonlinear Time Series Analysis. *Int. J. of Bifurcation and Chaos*, **5**(2), 349 – 358 (1995).
- [39] B. KERNIGHAN, D. RITCHIE. *Programmieren in C*. Carl Hanser Verlag, München, Wien (1983).
- [40] G. SCHNEIDER, R. ZÖLLER, J. HOLZFUSS. Adaptive Modellierung nichtlinearer technischer Prozesse und Systeme. In *Technische Anwendungen von Erkenntnissen der Nichtlinearen Dynamik*. VDI-Technologiezentrum Physikalische Technologien, Düsseldorf, S. 283 – 286 (1999).
- [41] J. ARNHOLD, P. GRASSBERGER, K. LEHNERTZ, C. ELGER. A Robust Method for Detecting Interdependences: Application to Intracranially Recorded EEG. *Physica D*, **134**(4), 419 – 430 (1999).
- [42] T. BUZUG, K. PAWELZIK, J. VON STAMM, G. PFISTER. Mutual information and global strange attractors in Taylor-Couette flow. *Physica D*, **72**, 343–350 (1994).
- [43] A. RÉNYI. *Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie*. Dt. Verl. d. Wissensch., Berlin (1972).
- [44] B. POMPE. Measuring statistical dependences in a time series. *J. Stat. Phys.*, **73**, 587 (1993).

- [45] H. BAUER. *Maß- und Integrationstheorie*. Walter de Gruyter, Berlin, New York (1992).
- [46] S. RESNICK. *Adventures in Stochastic Processes*. Birkhäuser, Boston, Basel, Berlin (1992).
- [47] W.-H. STEEB. *A handbook of terms used in chaos and quantum chaos*. BI-Wissenschafts-Verlag, Mannheim, Wien, Zürich (1983).
- [48] S. KULLBACK, R. LEIBLER. On Information and Sufficiency. *Ann. Math. Statist.*, **22**, 79 – 86 (1951).

Danksagung

Herrn Professor Dr. T. Tschudi danke ich für die Ermöglichung dieser Arbeit am Institut für Angewandte Physik der Technischen Universität Darmstadt. Seine Unterstützung und Förderung waren von großem Wert.

Dank schulde ich allen Mitgliedern der Arbeitsgruppe Nichtlineare Physik für die gute Arbeitsatmosphäre und die vielen interessanten Diskussionen.

Dabei gilt mein ganz besonderer Dank Herrn Dr. J. Holzfuss. Seine unermüdliche Bereitschaft zur Hilfe und seine stets neuen Anregungen haben diese Arbeit entscheidend vorangebracht.

Herrn Dr. M. Rüggeberg danke ich für die fachliche Unterstützung und seine Mitarbeit bei Fragestellungen bzgl. netzwerkfähiger Programme.

Meinen besonderen Dank verdienen auch Herr Professor Dr. W. Elsässer und die Mitglieder seiner Arbeitsgruppe, insbesondere Herr Dr. I. Fischer, Herr Dipl.-Phys. T. Heil und Herr Dr. T. Burkhard für ihre Unterstützung in Form von hilfreichen Diskussionen und der Bereitstellung ihrer Meßergebnisse.

Zu danken habe ich aber auch den Mitarbeitern der Unternehmen Schenck Pegasus GmbH (Herr Dr. V. Tews), Schenck RoTec GmbH (Herr Dipl.-Phys. A. Buschbeck), Schenck Vibro GmbH (Herr Dr. M. Weigel) und Schenck Process GmbH (Herr Dipl.-Ing R. Müller), sowie den ehemaligen Mitarbeitern der zentralen Forschungs- und Entwicklungsabteilung der Carl Schenck AG Herrn Dr. R. Zöller und Herrn Dr. F. Reine für die gute Zusammenarbeit und interessanten Anregungen.

Diese Arbeit wurde durch das BMBF / VDI (physikalische Technologien) im Rahmen des Projekts *Analyse von Maschinen- und Prozeßzuständen mit Methoden der Nichtlinearen Dynamik* (FKZ 13N7041/0) finanziell gefördert.

Lebenslauf

Olaf Göhrs

geboren am 19. April 1968 in Göttingen

1974 – 1978	Besuch der Grundschule Röddenberg in Osterode
1978 – 1980	Besuch der Orientierungsstufe Röddenberg in Osterode
1980 – 1987	Besuch des Gymnasiums Osterode
Juni 1987	Abitur
1987 – 1989	Zivildienst beim Rettungsdienst des DRK in Osterode
1989 – 1990	Studium der Chemie an der Universität Oldenburg
1990 – 1996	Studium der Physik an der Universität Oldenburg
Juni 1996	Diplom in Physik
1996 – 1997	Netzwerk-Administrator bei der Kraemer GmbH, Osterode
seit August 1997	Wissenschaftlicher Mitarbeiter am Institut für Angewandte Physik, Technische Universität Darmstadt